Developing Tools for Identifying and Analyzing Amtrak’s Most Delayed Trains

By: Aidan Lynott

Advised By: Jason Klusowski

June 2023

Submitted in partial fulfillment

of the requirements for the degree of

Bachelor of Science in Engineering

Department of Operations Research and Financial Engineering

Princeton University

I hereby declare that I am the sole author of this thesis.

I authorize Princeton University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Aidan Lynott

Aidan Lynott

I further authorize Princeton University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Aidan Lynott

Aidan Lynott

Abstract

A major problem with long-distance passenger train travel on Amtrak is the constant delays. While it makes sense that a train is more likely to arrive to a station late than early, as a train always has to wait at each stop until its scheduled departure time and thus cannot easily gain time, it makes less sense how delays of up to an hour are common, or even expected, on many Amtrak trains today. For example, Amtrak's train 5 had a mean arrival delay of 2 hours and 13 minutes and a median arrival delay of 1 hour and 26 minutes at its final station throughout 2021 and 2022. This is still clearly bad. But compared to other Amtrak trains, how bad actually is it?

While there has been analysis into the delay of Amtrak services, such as the Northeast Regional or the California Zephyr, there are no public resources available for identifying and ordering individual Amtrak trains – such as train 5, train 29, or train 319 – by their lateness. This thesis aims to identify which trains on Amtrak have the greatest mean and median delays by analyzing historical Amtrak delay data and listing Amtrak trains by their calculated statistics of net magnitude of delay and delay normalized by route length within two time frames between 2021 and 2022. Further, once the most delayed trains have been identified, the thesis creates tools that analyze why these trains are as delayed as they are by visualizing the distributions of their delays, using hypothesis testing to verify their lateness, and analyzing data about the delay causes from the Federal Railroad Administration. Finally, this thesis uses the created tools to analyze the delay of selected Amtrak trains to find which trains have delay that is possible for Amtrak to fix, and which trains are fated to be delayed by causes outside of Amtrak's control.

Acknowledgements

I've got a lot of people I want to thank, so let's get straight into it.

First, I'd like to thank my thesis advisor, Jason Klusowski. Throughout this thesis process, he helped me come up with ideas, validated that I was on the right track, and dealt with reading all of my long reports, including this one; thus, this thesis would not have been possible without him. In addition, I'd also like to thank the Princeton ORFE department for leading me through a unique and enriching program of study that would not have been possible anywhere else, as well as Princeton as a whole for being a place where I could not only prove what I'm capable of, but also have a home where I could continue to grow both personally and academically.

Next, my friends. I've got more friends that I'm thankful for than I can write here, but I'm still going to laud a few. To those who've been by my side for as long as I can remember: Ryan Huber and Alex Benfield; to the Hunterdon Central squad whose power continues to grow with every passing day: Vish Srikrishna, Troy Powell, AJ Bergamo, Austin Preiss, and Sam Nykorczuk; to one I've been able to count on for eight years, and hopefully for many more: Jess Stikons; to those I met at Princeton on day one that have made my four years here worth every second: Bryan Foo, Pablo Blanco Hinojosa, Keith Zhang, and Freddy Qui; to those who I met through Smash Bros, and have stuck with for everything else: Jack Isaac and Lucas Linzmeier; to Club Flag Football, the greatest organization on campus, and all the fantastic times I've had with all the incredible people I've met through it; and to those that without their help, this thesis in its current form would not have been possible: Lars Peterson, who gave me the suggestion to use the

Google Maps API to find the latitude and longitude of each Amtrak train station, and Jason “Zouba” Zou, who informed me of the Federal Railroad Administration’s published reports and delay metrics which made the second and third part of this thesis possible. I love and appreciate every one of you, and I look forward to seeing what great things you continue to achieve far into the future.

Finally, my family. Every single member of my family, whether distant or near, has been nothing but supportive of me for my entire life. I am beyond thankful for every single one of them, even as there are too many to name. However, I still feel obliged to send my thanks to those who have always been the closest; to my dog Chess, who, despite not being able to read this, I’m sure will appreciate the treats and pets I give him when I see him next; to my sisters, Emily and Lainey, who despite their youth have already achieved great things, while leaving no doubt that they will continue to achieve much more; and to Mom and Dad, who have always done everything they could to put me in the best position to succeed, no matter what obstacle stand in the way, so long as I try my best. I couldn’t be where I am today without your endless love and support, and I’m excited to see where it helps me reach as my life continues.

“No matter how ridiculous the odds may seem, within us resides the power to overcome these challenges and achieve something beautiful; that one day, we'll look back at where we started and be amazed by how far we've come.”

- Alexander “Technoblade”

Table of Contents

1. Introduction

On the night of March 9th, 2022, I was on a train heading through Utah. As a passenger on Amtrak's train 5, the California Zephyr, I was traveling from Chicago to California to visit a friend at U.C. Berkeley. And after a full day of appreciating the sites and the scenery, as the last person still in the observation car, I had a thought in my mind. *How late are we going to be?* I had heard bad things about the delays on long-distance Amtrak trains, and as we approached Salt Lake City, we had started to lose some time; we had been an hour late to the previous two stops. The train I had taken from New York to Chicago a few days earlier had done alright, ending up about twenty minutes late to its final station, but it had never fallen as far behind as we were now. Was my current train going to reach Berkeley on time, or at least in a reasonable time?

The train did end up being about one hour late, and although I wasn't happy with the result, I couldn't help but wonder: *How late was I, really? How late could I have been?* I went to the internet to see how train 5 typically performed, but a quick google search didn't give many answers; I only found from Amtrak's own site that the California Zephyr had an customer on-time percentage of 12% [11], a reference point that I already knew I had been on the wrong side of. But information about how late my individual train had been compared to other runs of train 5, or even how the delay of train 5 tended to compare to the delay of other Amtrak trains, was nowhere to be found. I knew that if I wanted to find out how late my train had been, whether compared to others of its own or to other trains entirely, I'd have to do some research and figure it out myself.

As I began searching for answers, however, I quickly became interested in more than just how late my train had been. Not only was my interest growing into how the delay of all Amtrak trains compared to each other, but I was also curious as to which Amtrak trains were the most delayed, and what caused these trains to be as late as they were. And although there was once again very little easily accessible information about individual trains, I did find some background information that immediately shed some light on the situation:

Amtrak trains are notoriously delayed, especially in regions outside of the Northeast where Amtrak does not own the tracks. Amtrak states that the majority of their trains' delay is caused by freight train interference, when host railroads prioritize the movement of freight trains instead of prioritizing passenger trains as they are legally obliged to [11]. According to Amtrak, freight train delay caused 900,000 minutes of delay in 2021 [11]. Meanwhile, most if not all of Amtrak's long-distance trains continue to fall far behind the 80% customer on-time performance metric that the Federal Railroad Administration has set as the standard [27], with services such as the Southwest Chief and the Sunset Limited falling below 10% [11].

While it was surprising to see just how prevalent the delays on Amtrak trains were, and how large of a factor that freight trains outside of Amtrak's control had on their delay, I couldn't help but feel that the details I had found still fell short of what I wanted. It seemed obvious that Amtrak's long-distance trains would be some of the worst performing; traveling a long distance gives plenty of opportunities for the trains to be delayed! But did any shorter distance trains perform relatively as poorly or more poorly than the long ones, perhaps by gaining more delay time for every mile they traveled? In addition, even though freight train interference was clearly a huge issue for Amtrak trains, especially since it was both a major cause of delay as well as an issue outside of Amtrak's control, certainly it was not the only factor causing trains to be delayed; perhaps some poor performing trains had smaller issues that were causing them to be late, covered up by the influence of freight train interference on Amtrak trains as a whole, that if fixed could cause a large improvement in the train's performance.

With these thoughts in my head, I knew that I was ready to begin my research. I had gotten the background information I needed, and with my personal train trip now in the back of my mind, I was ready to look into the delay of Amtrak's trains. And as I started my journey to get to the answers I desired, I knew the questions that I would have to solve: What are Amtrak's most delayed trains, in terms of both their delay at their destination and their delay per mile of travel? Why are these trains so delayed, and is there anything Amtrak can do to improve their performance?

2. Identifying Amtrak's Most Delayed Trains

The first goal for my research is to identify which individual Amtrak trains are the most delayed, as if I eventually want to analyze what makes the most delayed trains as late as they are, I first need to know what trains the most delayed trains are. In order to pursue this goal, I will look at the arrival delay data of Amtrak trains at their destination within the time frame of 2021 and 2022, as well as within the time frame of July 2021 to June 2022; while the exact reason for needing both the shorter and longer time period will make itself clear in the second part of the report, looking at the trains in both of these time frames will allow me to analyze what trains have been the most delayed in a recent and relevant time frame, while also allowing me to avoid any disruptions in the delay data that may have been caused by the onset of the COVID-19 pandemic in early 2020. In addition, in looking at the delay of each Amtrak train I'll specifically be looking at the delay of each individual Amtrak train number, where each train number specifies a train that goes from a set origin to a set destination with a set schedule, with each train number being scheduled to depart from its origin a maximum of once a day.

In order to identify which Amtrak trains are the most delayed, I will create a list ordering Amtrak trains by their mean and median arrival delays at their final station. I plan to look at each train's final station in order to account for all the delays, as well as the time saved, accumulated over the whole route. However, this does make it so the list is unlikely to be relevant towards the delay at the intermediate stops of each train, so the relation between such stops and the delay accumulated will thus have to be analyzed independently for any train for which it may be relevant.

Furthermore, creating a list that orders the trains by the net delay at their final stations also has the side effect of longer distance trains typically being more delayed, as by having a longer distance to travel the trains thus have more chances to be delayed along their routes. Therefore, I will also create a list ordering the trains by their mean and median delays at their destination normalized by the approximate distance they travel along their whole route. This list will create a more realistic representation of which Amtrak trains actually run the worst by showing which trains gain the most delay per mile traveled.

Once these lists are created, I will be able to identify which Amtrak trains are the most delayed, both by net delay and by delay per distance traveled, and thus be able to analyze what causes the delay of a select number of them in a later section of this thesis.

2.1: Examining What Has Already Been Done

Throughout my preliminary research for this thesis, I was unable to find any public sources that look at the delay of individual Amtrak trains at their destinations; this of course is the motivation for me to make a list of them and their delays myself. However, there is some public analysis available of the lateness of Amtrak train routes and services, done by Amtrak, the Bureau of Transportation Statistics, and the Federal Railroad Administration.

The primary statistic that's available online for getting a sense of the lateness of various routes is Customer On-Time Performance (OTP), which lists the percentage of customers that reach their scheduled stop on time. For example, on their website, Amtrak includes information on the customer OTP of the majority of their long-distance routes for the month of June 2022; this metric ranges from only 6.3% of customers reaching their destination on time on the Sunset Limited, Amtrak's route connecting Los Angeles and New Orleans, to 48% being on time on the Empire Builder, which connects Seattle and Chicago [11]. Meanwhile, the Bureau of Transportation Statistics's website shows the same statistic for the Northeast Corridor, on which 83% of passengers reached their destination on time in 2021 [4]. And with a little bit of further searching one can find customer OTP metrics for both Amtrak services and individual train numbers published quarterly by the Federal Railroad Administration in their "Intercity Passenger Rail Service Quality and Performance Reports" (reports that we will use extensively as we look to analyze the delay of trains later in this thesis) [14]. This customer OTP data can give us an idea of what train services are the most delayed; it's expected that the trains with the lowest customer OTP would also be the most delayed, as if customers aren't reaching their destination on time, then the trains can't be reaching their destinations on time.

However, customer on-time performance is not equivalent to delay, and one reason for that is that customer OTP doesn't say anything about how delayed a train tends to get once it's already not on time. For example, a train that's always one minute off its customers being on time would have a 0% customer on-time percentage, but other than that would still run pretty well, while a train that gets every passenger to its stop on time in the first half of its route, but is an hour late to every stop on its second half, would have around a 50% customer on-time percentage while running very poorly in its second half. In addition, another difference between customer OTP and delay is that, according to the U.S. Department of Transportation, a train can be slightly delayed but still be considered "on-time" [4]. For example, in the shortest case, a train trip of less than 250 miles is considered on-time if it's less than 10 minutes late at its endpoint, while in the longest case, a train trip of more than 551 miles is considered on-time if it's less than 30 minutes late [4]. Thus, while customer on-time performance may give us an idea of what trains tend to run poorly, it doesn't exactly tell us which trains are the most delayed.

Percentage of On-Time Customers by Route - June 2022		
Auto Train	Orlando – Washington, DC	**25%**
California Zephyr	Emeryville – Reno – Denver – Omaha – Chicago	**12%**
Capitol Limited	Chicago – Cleveland – Pittsburgh – Washington, DC	**30%**
Crescent	New Orleans – Atlanta – Charlotte – Washington, DC – New York	**35%**
Empire Builder	Portland/Seattle – Havre – St. Paul – Milwaukee – Chicago	**48%**
Silver Star	Miami – Tampa – Savannah – Raleigh – Washington, DC – New York	**15%**
Southwest Chief	Los Angeles – Flagstaff – Albuquerque – Newton – Chicago	**7.3%**
Sunset Limited	Los Angeles – Tucson – San Antonio – Houston – New Orleans	**6.3%**

Amtrak's publicly displayed Customer OTP metrics for their long-distance routes in June 2022.

Beyond the occasional mention of customer OTP on the website of Amtrak or a government entity associated with it, there are minimal mentions of how much specific Amtrak trains or services are delayed. Only the FRA's Performance Reports have any data surrounding the subject, as it includes metrics such as the total number of delay minutes each train number accumulated over a whole financial quarter, as well as the delay gained per 10-thousand miles traveled for most Amtrak services, but not individual train numbers [14]. But in addition to these metrics having properties that make them unsuitable for the lists I am trying to make – as total delay minutes is heavily influenced by how often a train runs and also doesn't incorporate any time the trains make up along their routes, and the delay gained per 10-thousand miles traveled for each service tells us little about the delay for individual trains – these metrics are also very hard to view and understand without processing their data first. Creating code to view these metrics in a simpler manner will be a large priority for the second part of this thesis, as we work to analyze why specific trains are late, but for now they do not help me in actually identifying the latest trains.

2.2: Understanding the Data Sources

2.2.1: Amtrak Status Maps Archive Database Train Delay Data

The primary data source I use for the identification of Amtrak's latest trains is the Amtrak Status Maps Archive Database [16]. The database, nicknamed "ASMAD" (and referred to as such throughout this thesis) by its creator, meteorologist Christopher Juckins, is a database containing the arrival times and delays, departure times and delays, and station data of nearly all Amtrak train runs dating back to July 4th, 2006. On the database's website, you can search for Amtrak train runs by train number, station, date range, delay, and even day(s) of the week. This allows me to download all the arrival delay data of any train at its destination throughout the relevant time periods.

The database gets its data directly from Amtrak, and thus the data it contains is as accurate as it can possibly be [18]. However, this does not mean it is perfect, as any mistakes from Amtrak can cause the data to be missing or inaccurate. For example, occasionally a run from an Amtrak train will not be tracked all the way through and thus will not make it into the database. One case of this comes from January 11, 2023's run of train 53, the Auto Train from Lorton, Virginia to Sanford, Florida; despite being in the database as having left Lorton 21 minutes early, this train's arrival at Sanford 37 hours later is nowhere to be found [22] [37]. If such an occurrence of missing data were to be common throughout the database, it could be a bad sign for my usage of the database as a whole, especially when I am looking to use the delay of trains at their destination. However, ASMAD has a page where you can see the percentage of available data for arrival delays for each train, so you can check if trains are missing a lot of the data at their destinations [24]. And luckily for my research, the trains all seem to be pretty well upheld; from 2021-2022, train 53 has 92% of its arrival delays in the system, while most other trains, such as train 5, have upwards of 95%. Thus, I view this inconsistency of the ASMAD data as notable, but not detrimental; these large samples of the train runs will provide us with more than enough information to be confident in our identifications of the worst performing trains.

Train 53 Average Delays by Station (minutes) Originating 01/01/2021 to 12/31/2022					
Station Code	**Station Name**	**Average Ar**	**Available Data***	**Average Dp**	**Available Data***
LOR	Lorton - Auto Train Only, VA			3	97%
SFA	Sanford - Auto Train Only, FL	46	92%		

Train 53 can be seen to have 92% of its arrival delay data available at its destination.

Another rare issue with Amtrak data, and thus the data in ASMAD, is that occasionally arrival times for trains are misinput, causing errors in their delay data. For example, according to ASMAD the earliest train ever was a 23 hour early train on May 31, 2007, as train 52, the Auto Train that goes from Florida to Virginia, left its origin at 3:53PM, only to time travel and reach its destination over six hours earlier on the same day at 9:38AM [21]. Looking at the data though, it is obvious what actually happened: the train arrived at 9:38AM the next morning, completing its journey 8 minutes late, and was misinput as having arrived the same day that it left. However, the fact that such a misinput could happen is incredibly concerning. How often does such a thing occur; is it often enough to mess up my research?

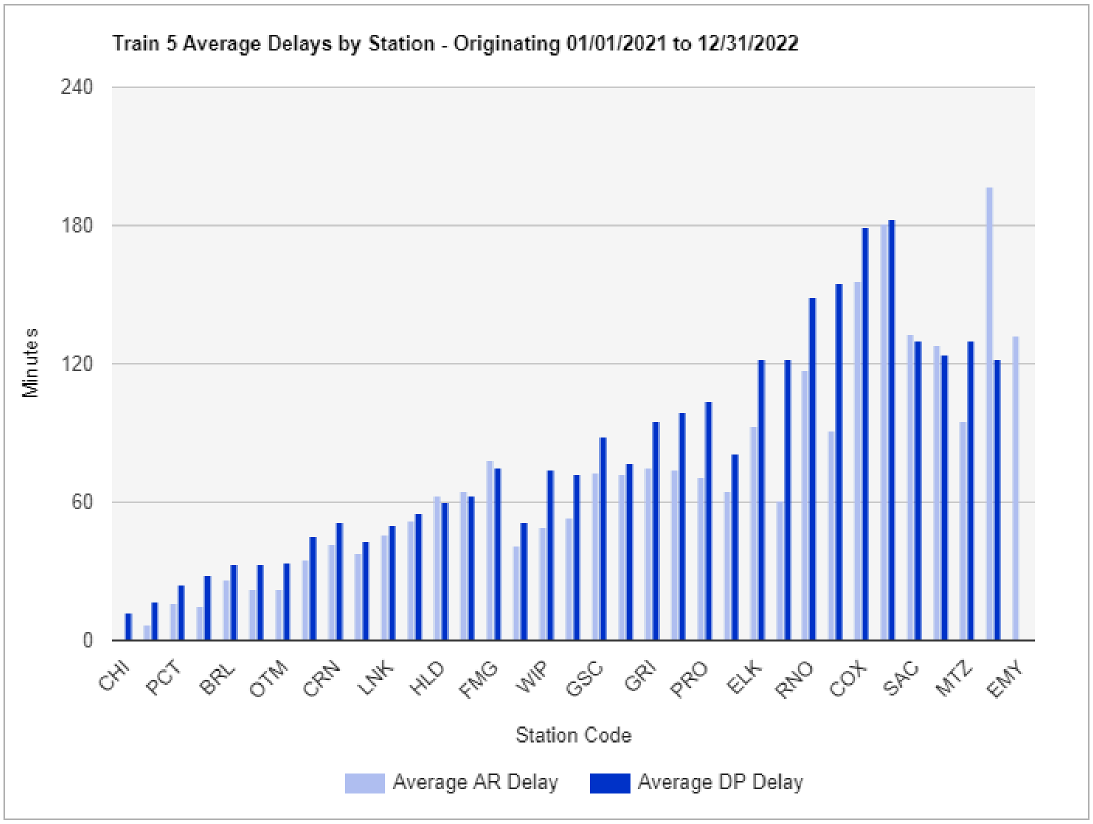
Luckily once again, the answer is not often. In the entire history of ASMAD's data, only 11 train runs have been input as being three or more hours early, with all of them being misinputs. Every one of these trains have easy explanations as to why they were "early," ranging from misinput AMs/PMs to having defaulted arrival times of midnight, and only one of them falls within the range of time I'm looking at. If I just ignore these super-early trains in my analysis, then I should have nothing to worry about.

But what about trains that aren't quite as early, but are still early? If you look at trains that have arrived more than two hours early, are all these misinputs too, or have some unforeseen circumstances actually caused these trains to speed ahead? As I looked through the data, I found that while many of these were clear misinputs (many of which had their true arrival time off by three hours, probably due to a time zone difference between the train and the inputter), some of them weren't, as sometimes trains did zoom through their route, skipping stops as they went to get where they needed to be.

Between 2021 and 2022, the earliest train I could find that was actually early was 2 hours and 9 minutes early; thus, as I download and analyze the train data I need, I will throw out any train run I find that finished earlier than this. However, I know that there will still be trains that make it into my data with incorrect early arrivals that any code I write won't be able to account for, as when the trains get less early, more accurate arrivals will be mixed in with the falsified ones. Thus, to avoid the effect of these potential misinputs, I will primarily concern myself with the median delays of trains, as the median delays will be less affected than the mean delays by any outliers caused by the incorrect data. However, since these misinputs do happen incredibly rarely, statistics relating to the mean delay should not be affected much; any effect may not even be noticeable at all (or may not exist for trains without any misinputs) and thus anything interesting we find using the mean delay of a given train should not be ignored.

2.2.2: Other Useful ASMAD Tools

In addition to the vast amount of train delay data that ASMAD contains, it also contains two useful tools that I make use of within this report. The first is a list of all Amtrak stations, allowing me to pinpoint the location of each train's origin and destination just through the stations' three-letter codes [15]. The second is a tool that allows you to graph the average arrival and departure delay of a train number at each of its stations over a set period of time [24]. Using this tool, you can see how the delay of a train tends to change as it goes from station to station, and perhaps identify where a train may be more likely to experience delays along its route. While this aspect of ASMAD will be more useful in the second part of the thesis as I analyze the reasons that certain trains are delayed, the graph and its delay data will still have an important use as I create the lists of the most delayed trains.



An example of such a graph, showing train 5's average arrival and departure delay at every station.

2.2.3: Google Maps Latitude and Longitude Data

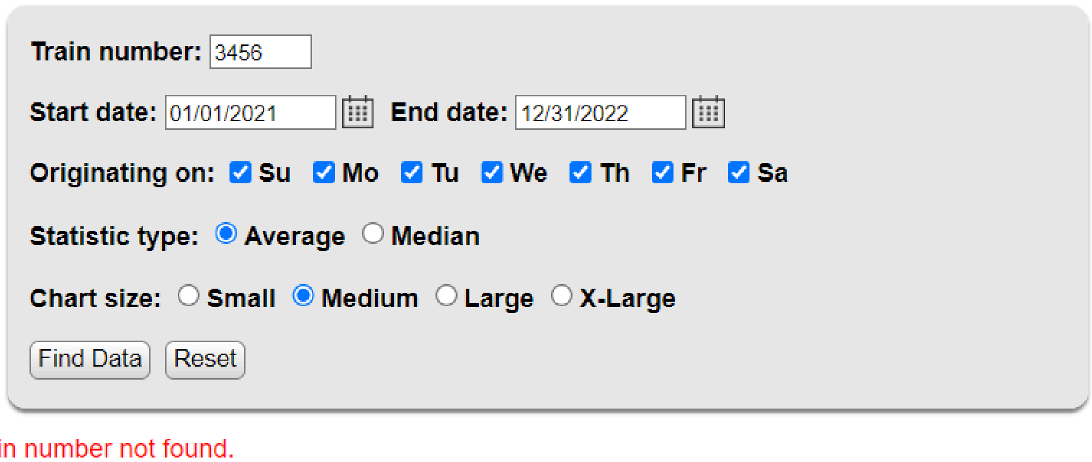
In addition to using the ASMAD data mentioned above, I will also use the Google Maps API, a tool which allows me to search Google Maps and get data from the results within my code. I will use this tool to search for and obtain the location data, and thus the latitude and longitude, of each station; this is essential for approximating the distance each train travels, as will be described later.

2.3: Writing the Code

2.3.1: Finding All Active Amtrak Trains

In order to identify which Amtrak trains were the most delayed in the time frames I was looking at – 2021 through 2022 and July 2021 through June 2022 – I first needed to know which train numbers had run within these time frames; this was necessary so that I could identify their destinations and thus download the arrival delay data at each train's destination over the course of the two years. However, as far as I can tell, there is no publicly available list of all currently active Amtrak trains, nor of Amtrak trains active over a certain time period. Thus, I needed to develop my own method to identify what Amtrak train numbers had run over a given period of time.

In my experimentations with the Amtrak Status Maps Archive Database, I recognized that when you searched for the data of a train's average delay at each station, if the train you were searching for had no data, it would return one of three messages: it would say "No Results" if you were searching for a real Amtrak train with no available data, "Train number not found" if you were searching for a train number that had never existed on Amtrak, and "Invalid train number" if you searched for something other than a train number, such as the word "data". Further, I found that the database made it impossible to search for a train number above 9999, and therefore I knew that 9999 had to be the upper bound on possible train numbers. Thus, I built a web scraper to search the web page of the station delay data for every train number from 1 to 9999; if the train's page did not have one of the three error messages, I would mark down that train number as existing and download the data on the page for future use. In this way I was able to build a list of all 525 active Amtrak trains within 2021 and 2022.



Searching for a train number that has never run.

2.3.2: Finding the Origins and Destinations of The Trains

Once I had downloaded this data for each train, I was then able to use it to obtain the origin and destination stations of each train. Since I had downloaded the average arrival and departure delay data for each train at each of their stations, and the data for the stations was typically in the order that a train visited its stations, for any given train all I had to do to determine its start and finish was mark the first station with data as its origin station and mark its last station with data as its destination station. However, for some of the trains the station data is slightly out of order – perhaps one station was placed before the true origin station, for example. Thus, I couldn't count on the first and last stations for each train being their origins and destinations. However, I further noticed that, even if the train's stations were out of order, only the origin station never had any arrival delay data and only the destination station never had any departure delay data, as you cannot arrive at the train's origin nor depart from the train's destination. Thus, to find each train's origin and destination, I marked the train's origin as the first station with no arrival delay data, and the train's destination as the last station with no departure delay data; I then made a list matching each train number to that train's origin and destination, making it easy for me both to see what the origin and destination of each train was marked as and to change either the origin or destination of any train if there was an error with the ones I had assigned to them.

Unfortunately, there were in fact errors with the origins and destinations of some of the trains. I quickly noticed that, depending on the train, the origin or destination given to me by my code could be one that was used by the train far less often than another origin or destination. For example, my code gave me train 821 as having a destination of Dallas when in the whole time frame of 2021 through 2022, Dallas had only been its final destination once. Instead, train 821 would always end at Fort Worth, except for the one time it went to Dallas. And this wasn't the only occurrence of my code telling me a train had a destination that wasn't the best one for me to analyze; with 525 trains to go through, how would I find them all?

After thinking about the problem, I decided that if a train arrived at a destination an average of once a week – so 104 times for the two-year time period and 52 times for the one-year time period – it probably had the right destination and at least had a destination I could analyze. Thus, I wrote some code to give me the number of times each train had arrived at its supposed destination, and manually checked the destination of trains that didn't arrive at their supposed destination often – as if a train had finished at one destination only a couple of times, then there may be another destination where it finished more often – by searching up their recent runs and seeing if they usually finished somewhere other than where my code had told me. Thus, I checked every train with a given "small" sample size to verify that they had the destinations I thought they should.

In the case that a train had multiple destinations it had previously stopped at, I looked at a couple of factors to decide which destination I should use for my analysis. My primary concern was frequency; if a train had one destination it went to rarely and one it went to more often, then I would use the more common destination. However, if multiple destinations had been commonly used as the destination of a train, I would try to prioritize the more recently used destination. An example of this comes from train 154, which in May 2022, switched from ending at New York to ending at Boston. Even though it had a much larger sample size of using New York as its destination, as it had done so for all of 2021, since it ended the sample by ending in Boston and since we are looking for ways to potentially improve the most delayed trains, I set its destination as Boston, figuring that looking at where the train terminates now would be more valuable for our analysis.

In the end, in the time frame of 2021 to 2022, I ended up changing the destination of 14 trains, while keeping a more uncommon but recently used destination for 10 trains. In the time frame of July 2021 to June 2022, only four trains ended up having to have their destination changed, with five keeping a more uncommon but recently used destination. A full list of the train destinations I changed can be seen as a commented list in my code.

Once I finished fixing the trains' destinations, I then had to see what errors I could find in the origins of the train. Even though I wasn't taking data from a train's origin, I still needed all the origins to be as accurate as possible so that later on, the distance I calculated of the train's route would be accurate. However, incorrect origins of trains were harder to find without checking every train, as I did not have a sample of how many times each train departed from an origin. I quickly thought of a solution, though; since most trains have a "sister train" that travels their same route, but backwards, I figured that I'd check the origin of every sister train to a train that I had to change the destination of, as if the destination of one train was wrong, it was likely that its sister train also had the wrong origin. With this method, I was able to find six trains that had the wrong origin and fix them, while also deciding to keep an eye out for inaccurate origins as I continued my research.

Finally, with the origins and the destinations of the trains made accurate to the best of my ability, I decided that I would make one adjustment to my lists of delayed trains once it was created; it would only show trains that reached their given final destination an average of at least once every two weeks in our time frame. This would assure that we would only see trains that had data to analyze, and we wouldn't just be able to say "the sample size is low" to explain why the trains were late. I want to be able to find reasons why trains tend to be late, and this requires a history of the train being late, as opposed to just a few unlucky runs; by only showing trains with a history of being late, I can analyze the trains that are actually concerning and may thus have the capabilities to be improved.

```
# Changed Destinations to most common destination
# 1105 Destination changed to WPR
# 1106 Destination changed to DEN
# 146 Destination changed to NYP
# 1567 Destination chcnged to OSD
# 167 Destination changed to NYP
# 1765 Destination changed to OSD
# 1777 Destination changed to OSD
# 1785 Destination changed to OSD
# 350 Destination changed to PNT
# 525 Destination changed to OKJ
# 536 Destination changed to SAC
# 711 Destination changed to OKJ
# 742 Destination changed to SAC
# 821 Destination changes to FTW

# Changed Origins to most common origin
# 1105 Origin changed to DEN
# 1106 Origin changed to WPR
# 1774 Origin changed to OSD
# 1784 Origin changed to OSD
# 822 Origin changed to FTW
```

The list of changed origins and destinations for train routes in 2021-2022, kept as a reference in my code.

2.3.3: Downloading and Organizing the Delay Data for Each Train

Once I had each train's origin and destination set, I was able to download each train's arrival delay data at their destination. Once again, I made a web scraper to search for and take this data from ASMAD; by searching for the arrival delay for each time a train number reached its destination within a given time frame, I was able to obtain the data that I needed.

With the delay data for each train at its destination, I had the data I needed to create a list and see which Amtrak trains had the worst performance by net delay by taking the mean and median of each train's arrival delay times and sorting the train numbers by these values.

2.3.4: Normalizing the Average Delays of Each Train by Distance

As mentioned earlier, however, making a list of each train's net mean and median delay was not my only goal, as I also wanted to take the lists of delays per train and normalize them by the length of each train's route. I expected that in the case of net delay, the longest routes would be the most delayed, as a longer route gives a train more chances to be delayed. Thus, to get a better sense of which trains performed poorly without regard to length, I wanted to see which trains were delayed the most often in the same amount of running time; essentially, I wanted to find the delay per mile traveled of each train.

The process to get such a list was simple – take my already existing list and divide each train's average and median delay by the length of the train's route. However, this would require a point of data I didn't yet have: how long is each train's route?

2.3.4.1: Deciding How to Get Each Train's Length

As far as I can tell, there is no resource, via Amtrak or via anywhere else on the internet, where you can put in a train's number and get the length of its route. Assuming I am correct in the non-existence of this resource, there are many reasons such a resource may not exist, with the prime reasons among them being the aforementioned inconsistency of various train's routes and the sheer number of Amtrak trains.

In addition, while the length of whole Amtrak services such as the Northeast Corridor (457 miles) [29] and the California Zephyr (2,438 miles) [1] are easily findable, it is very difficult to find the distance between two intermediate sections on a rail line, and thus difficult to find the rail length of a train that starts and stops at one or two intermediate stops in a service. And while there are old Amtrak timetables that list the milage of stops on their rail lines [17], in order to use the timetables effectively I would need to assign a rail line to all 525 train numbers I found before hoping that, if every train was part of a rail line that existed in the timetables, the origin and destination of every train would also be in these aged timetables. This is an ordeal that did not seem realistic, especially since I was not able to finalize the list of train numbers I would be looking at until very late in my research process.

Thus, instead of using the exact distance every train traveled along their routes to normalize their delay numbers, I decided instead to find and use an approximate distance that each train traveled for this purpose. And I decided that the best way to get this approximate distance would be to use the great circle distance between each train's origin station and destination station. However, in order to find the distance between the origin and destination of every train, I first needed to find the location of every Amtrak station.

2.3.4.2: Getting the Location of Every Amtrak Station

Finding the location of every Amtrak station was surprisingly easy. In contrast to finding every active train number, where there is no list easily available online, there is a list of all Amtrak stations, kept and maintained by ASMAD [15]. It includes "all Amtrak stations in the database” and lists the three-letter code of each station along with its city and state.

However, finding the distance between stations would require data that ASMAD did not have: the latitude and longitude of each station. And since I planned to use the great circle distance formula to find the distance between stations once I had the latitude and longitude of each station, I needed a way to get these coordinates.

I ended up using the Google Maps API to achieve this goal. With the Google Maps API, I could give a search term to Google Maps and receive the latitude and longitude of the first location result. For example, searching “ABE Aberdeen, MD” would give me the location, and thus the latitude and longitude, of the Aberdeen Maryland Amtrak Station.

This format did not work for every station though; “CUT Cut Bank, MT,” for example, would give me the city of Cut Bank, Montana instead of its train station. After testing a few possible formats, however, I was able to find one that worked for every station. In the case of Cut Bank, the search would be “Cut Bank, MT Amtrak Station”, and in the general case it would be “City, State Amtrak Station”. Using the Google Maps API with this method, I was able to make a list matching each three-letter station code to its latitude and longitude, which I could then later use to calculate the distance between any two stations.

2.3.4.3: Great Circle Distance Formula

I used the great circle distance formula to calculate the distance between stations; a picture of this formula can be seen below, obtained from igeocode.com [8].

$$d = \arccos(\sin\emptyset_{LAT1} \times \sin\emptyset_{LAT2} + \cos\emptyset_{LAT1} \times \cos\emptyset_{LAT2} \times \cos(\emptyset_{LON1} - \emptyset_{LON2})) \times R$$

$Note:$
$d = Distance\ between\ 2\ coordinates$
$R = Radius\ of\ Earth\ (approx.\ 6371.0090667KM\ mean\ radius\ defined\ by\ IUGG)$
$\emptyset_{LAT1} = Latitude\ of\ first\ coordinate\ in\ radian$
$\emptyset_{LAT2} = Latitude\ of\ second\ coordinate\ in\ radian$
$\emptyset_{LON1} = Longitude\ of\ first\ coordinate\ in\ radian$
$\emptyset_{LON2} = Longitude\ of\ second\ coordinate\ in\ radian$

An image of the Great Circle Distance Formula, obtained from igeocode.com.

2.3.4.4: Applying to Each Train and Making a List

At this point, it was easy to apply the coordinates I had found to finding the distance between stations. Now, when my code would calculate the average or median delay for a train at its destination, it would start by taking the three-letter codes for that train's origin and destination. Then, using these codes it would take the corresponding latitudes and longitudes from the train station list, before finding the distance between the origin and destination stations using the great circle distance formula. Finally, it would take the average or median delay it had found and divide it by the distance between the stations, before adding that delay, now in a format of "delay per miles," to the list of trains and their delays. And once all the trains and their delays per mile had been added to the list, the code would sort the list for viewing and analysis.

It is worth noting that, since we are using the shorter great circle distance instead of the longer rail length for each train, these delays per mile will be greater than what the trains' actual "delay per rail mile" would be. However, given that this property applies to every train, my delay per mile list should have approximately the same order as a hypothetically accurate delay per rail mile list. This consistency is the reason I am using the distance between origin and destination station for all trains, including those that have decisive rail mile amounts, such as the trains that go the full length of the California Zephyr – if I mixed trains with known rail mileages with those that didn't, the trains with accurate rail lengths would tend to have lower delay per mile values than their inaccurate counterparts, which would move the accurate rail length trains to the center of the delay per mile list and make the order of the final list more inaccurate and unreliable. And since we are more concerned with which trains are the most delayed, as opposed to how delayed those trains are, the order of the list being more accurate and reliable is much more important than having a perfectly accurate distance and delay per mile for some – but not all – of the trains.

2.4: Seeing the Results

With the code written, we are able to generate lists of Amtrak's most delayed trains in the time frame of 2021 and 2022, as well as the time frame of July 2021 through June 2022. One set of these lists is ordered by the trains' mean and median delays at their final destinations, while the other is ordered by the approximate delay gained per mile traveled for each train along their route. The tops and bottoms of these lists can be seen on the next pages.

With these lists, we are able to identify the most delayed Amtrak trains; the most delayed trains by net delay seem to always be some combination of trains 1, 2, 3, 4, 5, and 6, all of which are long-distance trains, while the most delayed trains by delay per mile bring out some other candidates, such as trains 499, 1014, 319, and 696, among others. As expected, our set of lists for the net mean and median delays show that the longest trains tended to have the most delay at their destination, as these trains have plenty of opportunities to get delayed along their routes. However, this trend does not appear in the list of trains ordered by delay per mile; while the long-distance trains still tend to appear in the bottom half of these lists, completely new trains appear at the very bottom. In the next part of this thesis, we will create tools to analyze some of these trains, before using these tools figure out what causes these trains to run so poorly.

The Least Delayed (Left) and Most Delayed (Right) Trains by Mean, 21-22

Mean (Minutes Delayed)

Number	Minutes	Sample Size
59	-16.992	527
548	-9.887	203
1584	-7.261	69
619	-7.0	278
609	-6.171	175
66	-5.65	526
1564	-5.371	70
536	-5.337	498
520	-4.726	201
412	-4.455	202
304	-4.449	303
671	-4.416	149
605	-4.143	504
531	-4.102	333
330	-4.015	460
292	-3.646	82

318	46.304	217
53	46.453	674
21	47.126	587
22	49.944	622
30	58.789	616
319	62.304	214
29	62.77	614
50	67.195	302
14	67.386	585
92	70.254	623
91	72.438	608
19	77.168	524
8	97.477	577
4	100.147	587
3	103.286	587
2	119.178	292
5	133.159	590
1	133.634	303
6	140.315	597

The Least Delayed (Left) and Most Delayed (Right) Trains by Median, 21-22

Median (Minutes Delayed)

Number	Minutes	Sample Size
59	-46.0	527
7	-19.5	570
58	-19.0	558
11	-15.0	576
66	-14.0	526
548	-13.0	203
304	-12.0	303
1564	-11.0	70
536	-11.0	498
64	-11.0	716
1584	-10.0	69
531	-10.0	333
364	-10.0	709
744	-9.0	169
520	-9.0	201
796	-9.0	252

89	22.0	602
516	23.5	96
99	29.5	216
14	35.0	585
195	36.0	133
50	36.5	302
92	37.0	623
30	40.0	616
8	41.0	577
319	46.5	214
29	49.0	614
91	50.0	608
19	53.0	524
4	54.0	587
3	60.0	587
2	75.5	292
1	86.0	303
5	86.0	590
6	94.0	597

The Least Delayed (Left) and Most Delayed (Right) Trains by Mean, July-June

Mean (Minutes Delayed)

Number	Minutes	Sample Size
59	-18.843	281
63	-18.134	335
295	-13.692	39
548	-11.107	112
609	-8.604	48
536	-5.904	250
619	-5.803	213
656	-5.542	48
291	-5.082	281
663	-4.819	72
330	-4.488	254
304	-4.449	303
1584	-4.37	27
531	-4.348	253
64	-4.302	358
250	-4.047	106

1014	55.625	32
318	57.487	39
53	59.98	354
92	64.702	356
14	65.619	315
50	67.655	148
30	67.983	343
91	69.084	358
319	72.632	38
29	79.003	342
4	89.709	320
19	90.414	285
8	91.473	319
3	102.66	321
2	125.836	146
6	129.346	324
5	138.808	318
1	141.737	152

The Least Delayed (Left) and Most Delayed (Right) Trains by Median, July-June

Median (Minutes Delayed)

Number	Minutes	Sample Size
59	-47.0	281
63	-25.0	335
11	-20.5	312
7	-20.0	311
58	-16.0	312
295	-15.0	39
548	-13.0	112
66	-13.0	178
291	-13.0	281
304	-12.0	303
536	-11.0	250
64	-11.0	358
307	-11.0	361
1564	-10.0	28
744	-10.0	108
531	-10.0	253

14	30.0	315
49	32.0	328
99	33.0	105
92	35.0	356
195	38.0	77
50	41.5	148
8	45.0	319
91	46.0	358
1014	48.0	32
4	48.0	320
30	48.0	343
319	53.5	38
3	55.0	321
29	61.5	342
19	62.0	285
1	75.5	152
2	79.5	146
6	81.0	324
5	94.0	318

Least Delayed (Left) and Most Delayed (Right) Trains Per Mile by Mean, 21-22

Mean (Minutes Delayed Per Mile)

Number	Minutes	Sample Size
548	-0.144	203
536	-0.078	498
412	-0.077	202
619	-0.076	278
520	-0.069	201
1584	-0.065	69
1562	-0.056	62
522	-0.05	452
330	-0.05	460
1564	-0.048	70
531	-0.046	333
605	-0.045	504
609	-0.04	175
540	-0.037	202
579	-0.031	288
607	-0.028	63

686	0.135	411
27	0.139	576
507	0.142	577
1777	0.146	97
319	0.151	214
470	0.152	464
694	0.155	204
529	0.157	61
696	0.158	174
476	0.174	461
464	0.182	217
1579	0.184	84
1590	0.196	84
1567	0.21	62
1784	0.25	95
516	0.25	96
499	0.257	217
494	0.314	444
1774	0.373	95

Least Delayed (Left) and Most Delayed (Right) Trains Per Mile by Median, 21-22

Median (Minutes Delayed per Mile)

Number	Minutes	Sample Size
548	-0.189	203
1562	-0.174	62
536	-0.16	498
520	-0.131	201
526	-0.116	142
522	-0.116	452
531	-0.112	333
720	-0.102	172
540	-0.102	202
744	-0.101	169
1564	-0.099	70
619	-0.098	278
1584	-0.09	69
746	-0.087	171
534	-0.087	292
412	-0.087	202

450	0.069	146
470	0.069	464
504	0.069	574
686	0.075	411
195	0.076	133
507	0.076	577
29	0.082	614
464	0.087	217
476	0.087	461
1774	0.087	95
1777	0.087	97
529	0.094	61
694	0.108	204
319	0.113	214
1579	0.116	84
1765	0.116	85
696	0.116	174
1567	0.188	62
516	0.194	96

Least Delayed (Left) and Most Delayed (Right) Trains Per Mile by Mean, July-June

Mean (Minutes Delayed Per Mile)

Number	Minutes	Sample Size
548	-0.161	112
536	-0.086	250
412	-0.069	135
295	-0.067	39
619	-0.063	213
63	-0.059	335
520	-0.056	134
609	-0.055	48
330	-0.055	254
531	-0.049	253
540	-0.042	113
522	-0.042	251
611	-0.041	66
605	-0.04	253
1584	-0.039	27
656	-0.036	48

30	0.114	343
504	0.117	352
693	0.12	101
488	0.121	105
686	0.124	247
507	0.129	358
497	0.129	46
416	0.132	33
29	0.133	342
318	0.139	39
694	0.148	101
1014	0.149	32
696	0.163	105
319	0.176	38
470	0.178	253
464	0.182	106
476	0.186	249
494	0.306	217
499	0.36	107

Least Delayed (Left) and Most Delayed (Right) Trains Per Mile by Median, July-June

Median (Minutes Delayed per Mile)

Number	Minutes	Sample Size
548	-0.189	112
1105	-0.181	36
536	-0.16	250
520	-0.124	134
1106	-0.121	36
522	-0.116	251
744	-0.112	108
531	-0.112	253
746	-0.102	107
720	-0.102	109
540	-0.102	113
526	-0.102	124
1564	-0.09	28
748	-0.089	109
412	-0.087	135
330	-0.086	254

684	0.058	239
50	0.058	148
503	0.065	356
470	0.069	253
504	0.069	352
507	0.069	358
99	0.071	105
686	0.075	247
195	0.08	77
30	0.081	343
464	0.087	106
497	0.095	46
29	0.103	342
476	0.104	249
694	0.116	101
696	0.116	105
1014	0.129	32
319	0.13	38
499	0.15	107

3. Creating Tools to Analyze Train Delay

With the lists of delayed trains completed, we can now pursue the second goal of this thesis: analyzing why the most delayed of these trains run so poorly and seeing if there is any way that Amtrak can improve their performance.

To do this analysis, we need to be able to understand the typical delay causes and delay durations of different trains. Thus, we will now create a variety of tools to help us do this. Some of these tools will be used to visualize and understand the delay times of trains, others will be used to compare the delay of trains to others with their same route, and others still will be used to organize and view what causes individual trains to be delayed. And when these tools are used in tandem, they should allow us to better comprehend how and why the trains we analyze with them get delayed.

Once all the tools are created, in this thesis's final part we'll apply them to some of the trains at the bottom of our lists to find out why they are as delayed as they are, and to see if improvements for these trains are possible.

3.1: Examining What Has Already Been Done

Once again, there has not been much analysis into what makes individual Amtrak trains late; there is no tool online where you can plug in a train number and get the reasons that that train most often gets delayed. However, on its website Amtrak offers one major cause: freight train interference. Amtrak states that the largest cause of their train delay is freight train interference, when the railroads that Amtrak travels on prioritize freight train movement on the tracks as opposed to passenger train movement [11].

Since Amtrak only owns about 3% of the rails that their trains travel on, primarily in the northeast [3], Amtrak trains often have to use railroads that Amtrak does not own. And even though these railroads are required to prioritize passenger trains over freight trains by federal law [3], many do not, which let freight trains cause over 890,000 minutes of delay to Amtrak trains in 2021 [3] – Amtrak even grades railroad companies on how much delay they cause to Amtrak passengers, with the two worst, Union Pacific and Norfolk Southern, receiving C+ and D- grades respectively in 2021 [3].

Should it be true that freight train delay is the primary delay cause for all Amtrak trains, it seems unlikely that Amtrak could substantially improve much of their trains' performances, as they cannot control what trains are prioritized on railroads they do not own. However, Amtrak does not specify what trains freight train interference tends to affect, and thus it may be the case that freight train interference is not an issue for all of their most delayed trains. Thus, in looking to analyze the most delayed Amtrak trains, we'll also be looking to find which trains aren't typically delayed by freight trains and thus are candidates for Amtrak to be able to improve, and which trains are heavily delayed by freight and thus are unlikely to be able to be fixed by Amtrak.

3.2: Examining What Has Already Happened

Beyond what Amtrak says, however, we have another resource if we want to find out what might cause trains to be delayed: the most delayed trains in ASMAD. Since ASMAD keeps record of nearly every train run in the past 10+ years, we can use it to find the train runs that were the most delayed in the database, and then research these trains to find out why they were so delayed.

Thus, I searched the database for all the trains in the database that were more than 10 hours delayed, and the database gave me exactly what I was looking for. The three latest trains that the database has data on are:

3. The June 26, 2015 train 5 run, which was delayed 36 hours [20].
2. The March 17, 2017 train 8 run, which was delayed 40 hours [23].
1. The January 23, 2021 train 3 run, which was delayed 42 hours [19].

Despite this valuable information, ASMAD had no information on the causes of these train delays. Thus, if I wanted to know why these trains had been delayed, I'd have to find auxiliary information online about why these trains were delayed from when they were delayed. And after searching various local news articles, train forum posts, historical weather reports, and tweets from those wondering what was going on as the trains were being delayed, I was able to piece together a story for each of them: June 26, 2015's train 5 run had been delayed due to historic flooding in Ottumwa, Iowa, which it had to travel through on its route from Chicago to Emeryville, California [5] [13] [28]; March 17, 2017's train 8 run had been delayed due to extreme weather that had caused multiple landslides near its starting point in the pacific northwest [25] [30]; January 23, 2021's train 3 run had been delayed due to a separate train derailment in Arizona along the route it was taking from Chicago to Los Angeles, as well as by extreme weather conditions in Arizona and New Mexico [9] [26] [35]. Together, these stories put together an idea of what tended to cause the very worst train delays; historically bad weather with a sprinkle of outside train interference as well. However, although these trains give an idea of what causes the very worst train delays, it remains to be seen if these factors – bad weather and outside train interference – are what cause typical train delays, or if something completely different is the cause of most normal train slowdowns.

3.3: Understanding the Data Sources

3.3.1: Federal Railroad Administration Performance Reports

After every fiscal quarter, the United States Department of Transportation's Federal Railroad Administration publishes an "Intercity Passenger Rail Service Quality and Performance Report" [14]. These reports contain many metrics concerning how Amtrak trains and services are performing, but the ones that are important for this report are their Customer On-Time Performance metrics, their Delays per 10K Train Miles metrics, and most importantly, their overall Delay Metrics.

Starting with the delay metrics, these metrics contain data, in the format of a table, about what specifically causes each Amtrak train to be delayed. Each row of the table gives the minutes delayed for a specific train, for a specific delay reason, on a specific host-railroad, and whether the delay was Amtrak's fault or the host-railroad's. For example, by looking at the first row of this data we can see that in the 3rd fiscal quarter of the fiscal year of 2022, Amtrak's 52 train was delayed by a Car Failure (A mechanical failure on any type of cars) that was Amtrak's fault on the Central Florida Rail Corridor for a total of 153 minutes.

A	B	C	D	E	F	G	H	I	J	K	L
Fiscal Year	Fiscal Quarter	Business Line Name	Service Name	Sub Service Name	Train Number	Host Railroad Name	Host Railroad Code	Delay Responsibility	Delay Code	Delay Minutes	Host Railroad Train Miles
2022	3	Long Distance	Auto Train	Auto Train	52	Central Florida Rail Corridor	FR	Amtrak	CAR	153	1237.1
2022	3	Long Distance	Auto Train	Auto Train	52	Central Florida Rail Corridor	FR	Amtrak	ENG	113	1237.1
2022	3	Long Distance	Auto Train	Auto Train	52	Central Florida Rail Corridor	FR	Amtrak	ITI	3170	1237.1
2022	3	Long Distance	Auto Train	Auto Train	52	Central Florida Rail Corridor	FR	Amtrak	OTH	5	1237.1
2022	3	Long Distance	Auto Train	Auto Train	52	Central Florida Rail Corridor	FR	Amtrak	SVS	91	1237.1
2022	3	Long Distance	Auto Train	Auto Train	52	Central Florida Rail Corridor	FR	Amtrak	SYS	44	1237.1
2022	3	Long Distance	Auto Train	Auto Train	52	Central Florida Rail Corridor	FR	Host Railroad	CTI	55	1237.1
2022	3	Long Distance	Auto Train	Auto Train	52	Central Florida Rail Corridor	FR	Host Railroad	DCS	18	1237.1
2022	3	Long Distance	Auto Train	Auto Train	52	Central Florida Rail Corridor	FR	Host Railroad	DSR	214	1237.1
2022	3	Long Distance	Auto Train	Auto Train	52	Central Florida Rail Corridor	FR	Host Railroad	PTI	4	1237.1

The beginning of the Delay Metrics data, for the 3rd fiscal quarter of 2022.

This table has just about everything we need to know to determine what tends to delay certain trains; the only knock on it is that more specific delay reasons may be lumped into more broad categories of delay, such as a delay caused by a landslide being included as a weather delay. But even with this issue, we can still use the metrics to find out how often specific trains have been delayed for specific reasons; this makes them the perfect resource for calculating what tends to cause delays for different Amtrak trains.

These delay metrics do have a problem, however; there aren't a lot of them. In fact, at the time of my research only four of them had been released. In 2021, Amtrak changed how they released their delay metrics to make them far more detailed for the public – these are the releases suitable for use in my research, and the first of them came from the fiscal quarter of July 2021 through September 2021 [14]. Further, Amtrak tends to take a long time to release their metrics; as of the time of my research the most recent release came from March 2022 through June 2022, having been published in November 2022 [14]. So, if I wanted to use the delay metrics to analyze the most delayed trains, I could only apply the data I got from the metrics to trains from July 2021 through June 2022, a one year window; this explains why I needed to make a list of the trains from July 2021 through June 2022 in addition to ordering them from their runs through all of 2021 and 2022.

The other two relevant metrics, Customer On-time Performance and Minutes Delayed per 10K Miles, also have data for every fiscal quarter from July 2021 to June 2022. The customer OTP metrics are pretty simple, listing the customer on-time performance of every Amtrak service and train number that ran in the data's fiscal quarter; this allows us to both compare the performance of certain trains to their associated service and to check that our delayed trains have bad customer on-time performance, as we'd expect, and see just how bad their customer on-time performance is. Meanwhile, the delay per 10K miles metrics list the delay minutes accumulated per 10-thousand rail-miles traveled for each Amtrak service; with these metrics I can order every Amtrak service by delay minutes per 10K miles, and use this listing to check whether or not my most delayed trains are a part of Amtrak services that have high or low delay minutes per 10K miles, while gaining an idea of how these trains compare to the rest of their service.

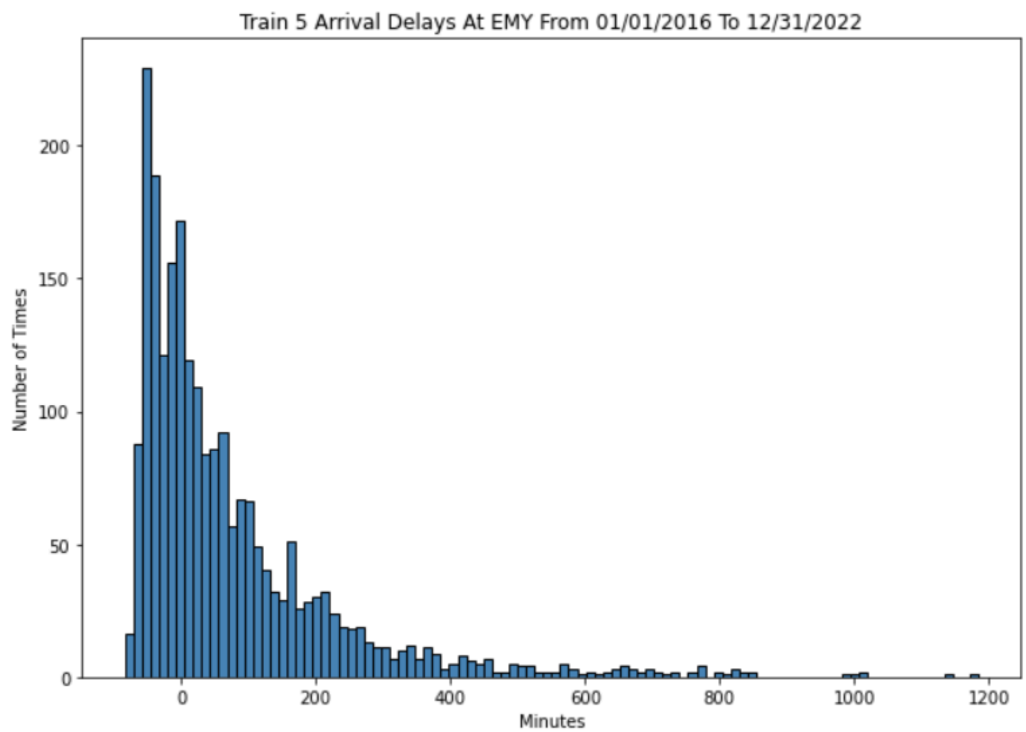
3.3.2: Amtrak Status Maps Archive Database

In addition to the metrics provided by the Intercity Passenger Rail Service Quality and Performance Reports, as we look to analyze why certain trains are delayed, we will once again use ASMAD's graphs of the average delay for each train at every station [24]. Touched upon briefly in the previous part of this report, these graphs will be essential in our analysis towards discovering if a train tends to get more delayed than usual between any two given stops; if we observe a greater increase in average delay between two stops on the route compared to the average delay increase between other stops, we can mark this place of a train's route as an area for improvement while also attempting to figure out what makes these parts of the trains' routes so much worse than the rest of the route.

3.4: Creating the Tools

3.4.1: Making a Histogram of a Train’s Arrival Delays

In wanting to be able to analyze the delay of any train, I decided that I wanted to be able to visualize each train’s delay at its destination. Thus, I wrote code that would do two things. First, given a train number, a station, and a time frame, it would go to the webpage of that train’s data on ASMAD and download the number of minutes of arrival delay the train had had every time it had arrived at that station within the time frame. Then, using these delay values, the code would make a histogram of that train’s arrival delays at that station for easy visualization and analysis, while also reporting the mean, median, and variance of the delays.




```
Mean 74.84946714031972
Median 22.0
Variance 25480.942259929205
```

An example of such a histogram for Amtrak train 5, between 2016 and 2022.

3.4.2: Organizing the Delay Metrics

Next, I wanted to be able to view the delay metrics for any given Amtrak train number. Thus, I wrote some code to analyze the delay metrics. I had three ideas for what I wanted the code to do:

1. If given a train number, I wanted to see who the responsibility for delays typically fell on for that train, what delay reasons were the most common, and who was typically responsible for what types of delays. This would allow me to see who and what causes certain trains to be delayed, and where fixes could be made.
2. If given a service name, I wanted to see who the responsibility for delays typically fell on for that service, what delay reasons were the most common, and who was typically responsible for what types of delays. This would allow me to compare a train number to other trains of the same service; if one train on a service tended to be later than the others, and if there was a delay reason that was more common on that train than on the rest of the service, then that delay reason may be a place where the more delayed train could realistically improve.
3. If given a train number and a delay type, I wanted to see who tended to be responsible for that delay type, and if the host-railroad was ever responsible, what railroad the host-railroad was. This would allow me to see which railroads were causing the most delay, specifically allowing me to see which railroads contribute the most to freight train interference.

Before long, I had the code written up. Some examples of the output of this code can be seen on the next page, looking at Amtrak train 5 and the rest of the California Zephyr. Notice that within this data, each Delay Reason is only mapped to one Delay Responsibility; this is a property of Amtrak's delay reasons and will always be the case.

You may also notice that the data for Amtrak Train 5 is quite similar to the data for the whole California Zephyr. In this case, this is to be expected, as train 5 makes up half of the California Zephyr's trips (with the other half belonging to train 6). However, this similarity may not hold for more busy Amtrak services, such as the Northeast Regional.

Delay Responsibility for Amtrak Train 5, Quarter 4 2021 - Quarter 3 2022

Delay Responsibility	Delay Minutes	Percentage
Host Railroad	102872	65.687577
Amtrak	30006	19.159941
Neither	23730	15.152483

Delay Reasons & Responsibility for Amtrak Train 5, Q4 2021 - Q3 2022

Delay Responsibility	Delay Reason	Delay Minutes	Total Percentage	Group Percentage
Amtrak	Crew & System	8269	5.280062	27.557822
	Servicing	5843	3.730972	19.472772
	Passenger Related	3889	2.483270	12.960741
	Miscellaneous Delays	3549	2.266168	11.827634
	Locomotive Failure	3080	1.966694	10.264614
	Car Failure	2648	1.690846	8.824902
	Passenger Related - Disabilities	1518	0.969299	5.058988
	Hold for Connection	694	0.443145	2.312871
	Injury Delay	476	0.303944	1.586349
	Initial Terminal Delay	40	0.025541	0.133307
Host Railroad	Freight Train Interference	49918	31.874489	48.524380
	Slow Order Delays	21605	13.795592	21.001828
	Passenger Train Interference	10338	6.601195	10.049382
	Signal Delays	9768	6.237229	9.495295
	Routing	6985	4.460181	6.789991
	Maintenance of Way	3180	2.030548	3.091220
	Commuter Train Interference	621	0.396531	0.603663
	Detour	457	0.291811	0.444241
Neither	Unused Recovery Time	10144	6.477319	42.747577
	Weather-Related	8083	5.161294	34.062368
	Police-Related	1977	1.262388	8.331226
	Trespassers	1646	1.051032	6.936367
	Debris	1300	0.830098	5.478298
	Drawbridge Openings	580	0.370351	2.444164

Delay Responsibility for the California Zephyr, Quarter 4 2021 - Quarter 3 2022

Delay Responsibility	Delay Minutes	Percentage
Host Railroad	207036	65.950784
Amtrak	61501	19.590985
Neither	45388	14.458230

Delay Reasons & Responsibility for California Zephyr, Q4 2021 - Q3 2022

Delay Responsibility	Delay Reason	Delay Minutes	Total Percentage	Group Percentage
Amtrak	Crew & System	14559	4.637732	23.672786
	Servicing	13197	4.203870	21.458188
	Passenger Related	8161	2.599666	13.269703
	Locomotive Failure	6802	2.166760	11.059983
	Miscellaneous Delays	6522	2.077566	10.604706
	Car Failure	4292	1.367206	6.978748
	Passenger Related - Disabilities	3317	1.056622	5.393408
	Initial Terminal Delay	2451	0.780760	3.985301
	Injury Delay	1254	0.399458	2.038991
	Hold for Connection	934	0.297523	1.518674
	Cab Car Failure	12	0.003823	0.019512
Host Railroad	Freight Train Interference	100017	31.860158	48.308990
	Slow Order Delays	44135	14.059091	21.317549
	Signal Delays	19497	6.210719	9.417203
	Passenger Train Interference	19254	6.133312	9.299832
	Routing	14229	4.532611	6.872718
	Maintenance of Way	7401	2.357569	3.574741
	Commuter Train Interference	2019	0.643147	0.975193
	Detour	484	0.154177	0.233776
Neither	Unused Recovery Time	17735	5.649439	39.074205
	Weather-Related	15968	5.086565	35.181105
	Trespassers	3969	1.264315	8.744602
	Police-Related	3782	1.204746	8.332599
	Debris	2586	0.823764	5.697541
	Drawbridge Openings	1348	0.429402	2.969948

Freight-Train Interference by Railroad for Amtrak Train 5

Host Railroad Name	Delay Responsibility	Delay Minutes	Host Railroad Train Miles	Percentage	Per 100 Miles
Union Pacific	Host Railroad	38207	452862.6	76.539525	8.436775
BNSF Railway Company	Host Railroad	11711	339303.0	23.460475	3.451487

Freight-Train Interference by Railroad for the California Zephyr

Host Railroad Name	Delay Responsibility	Delay Minutes	Host Railroad Train Miles	Percentage	Per 100 Miles
Union Pacific	Host Railroad	77598	928373.6	77.584811	8.358488
BNSF Railway Company	Host Railroad	22419	680663.9	22.415189	3.293696

3.4.3: On-Time Performance and Official Delay Per Mile

Keeping my attention on the Intercity Passenger Rail Service Quality and Performance Reports, I now turned to the two other useful metrics I would use to verify the analysis I was making: Customer On-Time Performance and Delay Per 10K Miles.

I started by looking at customer on-time performance. I figured that if I had a train my data said was consistently delayed, I could fact-check this constant lateness by checking that train's customer on-time performance, as if a train was often delayed, you would expect to see a low customer on-time performance. Conversely, if a train that appeared to be often delayed actually had high customer on-time performance, then perhaps the delay of the train isn't an issue or had been fixed in the latest fiscal quarter, or maybe something else was going on. Before long, I had code that could take either a train number or an Amtrak service and output its customer on-time performance. This would also allow for the comparison of a given train to the other trains of its service to see if that train was worse than the rest of its service at getting its riders to their destinations on time. The only negative of the customer on-time performance data is that since it comes as a proportion, I cannot add together the customer on-time performance of each fiscal quarter of data to get an all-encompassing customer on-time performance percentage. Thus, I will only be looking at the 2022 Quarter 3 customer on-time performance data, to see if potentially late trains are or aren't getting their customers where they need to be in the most recent time frame. Below, see two examples of the customer on-time performance output of my code, one for train 696 of the Downeaster, and one for the whole aforementioned service – notice that since train 696 has a lower customer on-time performance proportion than the rest of its service, it is probably more delayed than is typical for its service.

Train Number	C OTP
696	0.590683

Service	C OTP
Downeaster	0.846586

Customer On-Time Percentage for train 696 and the Downeaster service.

Next, I decided to look at the Delay Per 10K Miles metrics. It's important to remember that because of how I made my delay per mile metrics, by using the direct distance instead of the rail distance between train origins and destinations, you cannot directly compare the delay per mile values I have for a given train to the exact delay per mile values Amtrak shows for a given service; my values, due to having a smaller distance, should tend to be larger than Amtrak's. However, Amtrak makes up for this by allowing me to order their services by delay per 10K miles. Thus, if I have a train that my data says is often delayed, I can compare that train to where its service appears on the list; if its service is one of the most delayed services on the list, then I know I'm doing something right. And if its service is not one of the most delayed services, then either the train I'm looking at is really bad for its service, or there's a reason that the train I'm looking at is not usually delayed as badly as it initially seems. Either way, Amtrak's delay per 10K miles metrics is an excellent resource for me to utilize in my analysis of highly delayed trains, as it provides extra context to the delayed trains I find and can suggest why these trains may or may not be as bad as they seem.

Below, you can see the output of my code for working with these metrics. First, see how the code can output the delay per 10K miles for one specific service, and after see the whole list of services ordered by delay per 10K miles.

The Official Delay Per Mile of the California Zephyr

Service Name	Delay Minutes	Train Miles	DelayPerMile
California Zephyr	268537	3224939.8	0.083269

A List of Amtrak Services Sorted by Delay Per Mile

Service Name	Delay Minutes	Train Miles	DelayPerMile
Keystone	28662	2084937.4	0.013747
Acela Express	93098	4237511.6	0.021970
Northeast Regional	344776	9269748.8	0.037194
Empire	158256	3337606.4	0.047416
Hiawatha	38964	810443.2	0.048077
Capitol Corridor	89294	1824185.2	0.048950
Piedmont	38781	741251.0	0.052318
Palmetto	63586	1154647.8	0.055070
Vermonter	46399	832289.2	0.055749
Carolinian	56955	1013660.6	0.056187
Downeaster	58559	1020908.6	0.057360
Pennsylvanian	37916	640900.0	0.059161
Silver Meteor	66207	1103500.4	0.059997
Illinois	190453	2955062.0	0.064450
Empire Builder	213241	3277744.0	0.065057
Southwest Chief	196082	2982582.2	0.065742
City Of New Orleans	75733	1145069.0	0.066138
San Joaquins	176796	2634243.4	0.067115
Silver Star	154153	2186601.2	0.070499
Cardinal	50755	700342.2	0.072472
Pacific Surfliner	208987	2561142.0	0.081599
Missouri	44139	537137.0	0.082175
California Zephyr	268537	3224939.8	0.083269
Michigan	175410	1938202.8	0.090501
Cascades	102800	1131908.8	0.090820
Lake Shore Ltd	137508	1513592.8	0.090849
Crescent	158501	1691680.0	0.093694
Coast Starlight	164524	1754900.6	0.093751
Heartland Flyer	28556	297406.8	0.096017
Auto Train	128241	1232375.2	0.104060
Texas Eagle	194168	1791838.2	0.108362
Lincoln / Missouri	9754	86586.4	0.112650
Capitol Ltd	120981	1073830.4	0.112663
Sunset Ltd	155512	1199037.6	0.129697

3.4.4: Verifying Train Comparisons with Hypothesis Testing

Next, I wanted a way to verify that, given two trains and the delays they had over a given time frame, there was a statistically significant difference in the average delays of the trains. Obviously, over any time frame we look at, one train will have less delay than another, indicating that it ran better over that period. But what if the trains had been able to run many, many times over the same conditions of the time frame; how likely is it that the better train would have stayed with a lower average delay? In other words, can we verify that one train actually ran better than another, and that the difference in their delays wasn't just due to random chance? Knowing such a thing could help Amtrak choose which trains to focus on improving; should they have to choose between a train that looks worse but is harder to fix, and a train that looks better but is easier to fix, knowing whether or not there was a significant difference in the delays of the trains can help Amtrak choose which train to work on.

Therefore, in order to see if there’s a statistically significant difference in the average delays of any two trains over a time frame, I decided to write code that would use hypothesis testing. With hypothesis testing, we can analyze the likelihood that one train’s expected delay is greater or less than another’s during the time frame we’re looking at, given the runs we have from the trains in the time period.

At first glance, this may seem a bit odd to do, as hypothesis testing is used with sample data to verify a relationship with a population statistic. But aren’t we already looking at the population data for the train delays; why are we trying to verify the mean of the delays for a time frame when we already have it?

There are two reasons that remove this problem, however, with the second reason being more subtle. First, we don't actually have the population data for the trains to begin with; as established earlier, the database we're using, ASMAD, is missing some of the train data, and thus we're only looking at a sample of the arrival data for most trains. And second, we're not trying to find the population mean of the delays of the trains that ran over a given time frame, but what the population mean of delays of the trains would be had they been able to run many more times than they had under the conditions of time frame; this allows us to account for the variance of the runs to see which trains definitively performed better or worse than others. Thus, if we treat each arrival delay of the trains as a randomly sampled delay from all the possibilities in the time frame, we can use hypothesis testing to verify that one train, in terms of average delay, did perform significantly better or worse than another train.

Therefore, I created the code that, when given two trains, destinations for each train, and a time frame, would download the delay of the trains at the stations and run a 2-sample t-test. This would test the difference in the average arrival delays of the two trains over the time frame and return a p-value, representing the odds of obtaining a difference in average train delays greater than or equal to the difference that we obtained if the two expected delays were actually the same. Therefore, if the p-value were sufficiently small, we'd know that there was a significant difference in the average delays of the trains, and that one train did run significantly worse than the other. Below, you can see an example of this for train 1, ending at Los Angeles, and train 2, ending at New Orleans, throughout all of 2021.

```
Train #1: Train 1
Destination #1: LAX
Train #1 Average Delay: 107.18181818181819 Minutes
Train #2: Train 2
Destination #2: NOL
Train #2 Average Delay: 93.93877551020408 Minutes
Time Frame: 01/01/2021 - 12/31/2021
Test Statistic: 0.8567378590997778
P-value: 0.39232301789503565
If the p-value is sufficiently small,
 train #1 is significantly more delayed than train #2 on average.
```

Comparing the delays of train 1 and train 2 in 2021 using Hypothesis Testing.

Furthermore, I also wanted to be able to compare the delay of the same train in different time frames, thus verifying a significant improvement or deterioration in the train's performance. Again, by random chance it is possible (and nearly guaranteed) that a train's mean and median arrival delays will be different across two different time frames, even if there is no change to how the train is being run. However, this does not mean that the time frame with a lower mean and median delay would actually perform better in the long run; with no actual changes to the train, random variance in the runs may have caused it to appear to run better, when in reality if each train could have ran many more times in the time frame's conditions it actually would have shown to run worse.

Thus, I wrote some more code to do the same thing as above, but instead for the same train at the same destination for two different time frames to verify whether the train's performance had gotten significantly better or worse between the time frames. Below, see an example of this for train 5, comparing its delays between 2021 and 2022.

```
Train #: Train 5
Destination: EMY
Time Frame 1: 01/01/2021 - 12/31/2021
Time Frame #1 Average Delay: 93.4779411764706 Minutes
Time Frame 2: 01/01/2022 - 12/31/2022
Time Frame #2 Average Delay: 167.1006289308176 Minutes
Test Statistic: -5.1309741388185115
P-value: 3.925859423456939e-07
If the p-value is sufficiently small,
 time frame #2 is significantly more delayed than time frame #1 on average.
```

Comparing the delays of train 5 in 2021 vs. in 2022 using Hypothesis Testing.

Finally, note that I only made these tools to analyze net delay, and not delay per mile. This is due to the inherent inaccuracies in my delay per mile metric since it is created using an approximate distance. The purpose of the delay per mile metrics is to give a general order for the worst performing trains in order to identify them, and not to be an exact number to be used in analyzing the trains; therefore it doesn't make sense to include the delay per mile values in my hypothesis testing, as you don't want to verify values that are already inaccurate, and thus you can only use net delay with these tools.

3.4.5: Tools for Analyzing Delays of Trains with The Same Route

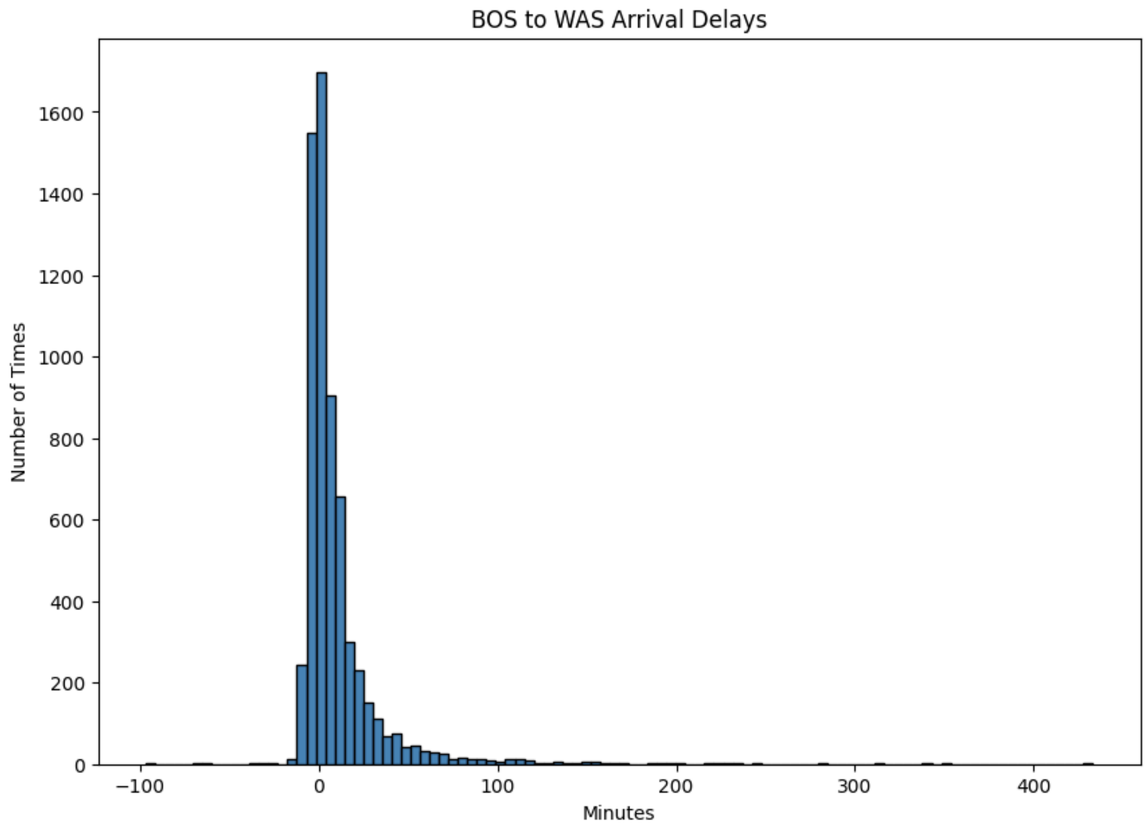
Finally, I decided to create some extra tools for analyzing delays; while I do not use the upcoming tools in my analysis of poor performing trains below, they could be useful to one who wishes to pursue further research in this area and/or analyze more than just the individual trains and their individual delays.

First, I decided that I wanted to be able to order and analyze trains of the same exact route based on their net delays and delays per mile traveled. Thus, I modified my code for making the lists of trains ordered by their delay statistics so that, when given an origin and destination, the list would only include trains with that origin and that destination. This would allow one to specify their analysis of trains to one specific section or route of Amtrak travel; an example of this can be seen below for trains traveling from Boston, Massachusetts to Washington, D.C. in 2021 and 2022.

Mean (Minutes Delayed)			Median (Minutes Delayed)		
Number	Minutes	Sample Size	Number	Minutes	Sample Size
2151	2.0	237	149	-7.0	74
2249	2.723	112	163	-3.0	81
149	2.73	74	2249	-3.0	112
2153	4.342	284	2257	-1.0	102
2257	4.912	102	2259	-1.0	105
139	5.97	99	2151	-1.0	237
173	6.309	501	139	0.0	99
2167	7.145	255	2167	0.0	255
2159	7.322	441	173	0.0	501
2259	7.324	105	2253	1.0	219
163	8.148	81	2153	1.0	284
2165	9.124	105	2159	2.0	441
2169	9.217	438	137	2.0	499
2155	9.333	501	2255	3.0	104
2173	9.539	440	2165	3.0	105
2253	9.726	219	2169	3.0	438
2251	9.983	116	2173	3.0	440
2255	10.183	104	175	3.0	449
137	10.246	499	135	4.0	219
165	11.65	103	2155	4.0	501
2163	12.587	499	165	5.0	103
175	12.646	449	2251	5.5	116
135	13.968	219	2163	7.0	499
177	14.659	126	161	8.0	190
161	20.258	190	177	9.5	126

Mean and median arrival delays for trains that traveled from Boston to D.C., 2021-2022.

Along the same line of thinking, I also decided to modify my histogram code for the same purpose; given an origin and destination, the code will make a histogram of the delay data for all trains with said origin and destination, as it does below for all the trains that traveled from Boston to Washington, D.C. in 2021 and 2022.




```
Mean 9.275023681717714
Median 2.0
Variance 598.8141630480656
```

Histogram of arrival delays for trains that traveled from Boston to Washington D.C., 2021-2022.

Next, I decided I wanted to make a tool to compare how often a train is less or more delayed than another train; this would allow for the comparison of net delay values for two trains with the same or similar routes to see which performs better. Therefore, given two trains, the tool says what percentage of the time each train has less delay than the other. One major perk of this tool is that it would achieve the same comparison for any two sources of delay data, whether coming from a train, a plane, or any other source. The code is also written to run efficiently for large amounts of delay data – by sorting the delays, going through each delay list at the same time, and counting the previous number of smaller delays at each new delay value, the code is able to avoid comparing every single delay number from each source and thus runs in loglinear time instead of quadratic time – and therefore it would work well if used for further analysis with larger amounts of delay data than I have analyzed in this report. Regardless of how and with what it is used, however, I believe that it is a useful tool that could find applications wherever there is delay data to be analyzed.

```
Train 3 has less delay than Train 4: 52.41620691356448% of the time!

Train 3 has more delay than Train 4: 47.252364548174675% of the time!

Train 3 has equal delay to Train 4: 0.33142853826084184% of the time!
```

Comparing the delays of trains 3 and 4, which connect Los Angeles and Chicago, in 2021 and 2022.

4. Analyzing Amtrak's Most Delayed Trains

Using the tools created and described above, I now plan to analyze the most delayed trains that we have identified in order to find out why they are so delayed, and if there is any way that Amtrak can fix their delay. I will start by analyzing the five most delayed trains in the list of trains ordered by median delay per mile from July 2021 to June 2022; this date range allows us to analyze trains that are in a time frame for which Amtrak's delay metrics are available, while using median delay per mile as our metric allows us to analyze trains without having to worry about them being often delayed because of outliers or often delayed just because they are really long trains. After I analyze these five delayed trains, I will analyze some other interesting trains on the list based on my own discretion.

4.1: The Five Worst Trains by Median Delay Per Mile, July 2021 Through June 2022

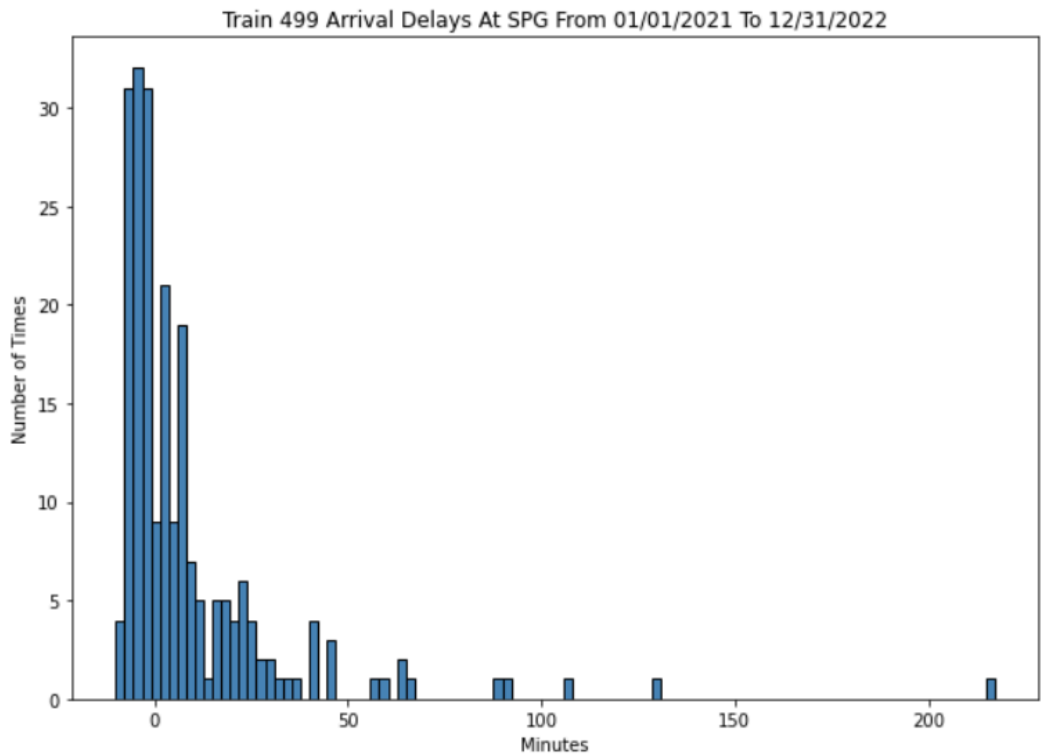
4.1.1: Train 499

Of all 433 train numbers that made at least 26 runs between July 2021 and June 2022, train 499 has the highest mean and median delay per mile on our list. Clearly, this train is bad. Just how bad is it?

First, I want to get a bit of context for train 499. Train 499 is a train on Amtrak's Valley Flyer, a service that goes from Greenfield, Massachusetts to New Haven, Connecticut. Train 499 does not follow this whole route, however, as instead of continuing all the way to New Haven, it instead stops at Springfield, Massachusetts, giving us only about a 33-mile distance from our start to our end. This makes train 499 a pretty short train, with its total scheduled run time being only just over an hour.

Next, I want to see how bad train 499 was over our entire two-year period, through all of 2021 and 2022. If train 499 is as bad as we expect it to be, then it should be near if not at the bottom of both of our time frames, especially because the time frame we've already looked at makes up half the sample size of our longer time frame. And when we look at the same lists for the full years of 2021 and 2022, train 499… falls over 20 spots away from the bottom in the list of median delay-per-mile. Maybe train 499 isn't as bad as we thought it was? Now, it's still bottom three in 2021-2022's mean delay-per mile, but that's very prone to outliers. Does train 499 have significant outliers?

Looking at the distribution of train 499's delays at its destination for our whole two-year period, we can see that it does, having three delays of over 100 minutes and one over 200.




```
Mean 8.56221198156682
Median 2.0
Variance 597.4811527108242
```

However, the statistics given to us with this graph paint an even clearer picture. Train 499, in all of 2021 and 2022, has a median arrival delay at its final destination of… 2 minutes? How can a train that arrives 2 minutes late or better 50% of the time be the worst train in the country, or at least one of the worst trains on my list?

The answer comes in train 499's extremely short distance. Being a 30-mile-long trip from start to finish, any delays are going to be very apparent in a delay per mile format, even the 5-minute arrival delay that was its median between July and June that puts it at the bottom of the delay per mile list. However, it is hard to say that a train that arrives five minutes late on a normal day is overly bad. Apart from the occasional large mishap, the train seems to run alright judging by its delay distribution; it's not great by any means, but to call it one of the worst seems a bit much. I'm torn here; is there another metric we can look at to see if train 499 urgently needs fixing?

Well, what about the train's customer on-time performance? If the majority of the train's customers reach their destination on-time, then it probably isn't worth the resources to full-focus on improving the performance of train 499. Is the train's customer on-time performance as bad as we'd expect from the last place train on our list?

Train Number	C OTP
499	0.833333

83% of train 499's customers reached their destination on time from April 2022 through June 2022. Is it perfect? No, but it's pretty good, and even meets the FRA's 80% customer on-time performance standard. And putting the time and resources into improving a train that already has 83% of its customers arriving on time is certainly not worth it, especially considering both the short distance of the train and the worse trains we will see later. Therefore, despite being the lowest, most delayed train on our list, train 499 and its delay is not an issue that Amtrak should prioritize trying to improve.

I do want to mention two things before we move on from train 499 though. First of all, the elephant in the room; how can my list of delayed trains be accurate and good to use when the lowest train on it isn't even that bad? Isn't it supposed to give us the worst trains? And to that the answer is, like all the other data sources I am using in this report, the list is only truly useful within the context that surrounds it. As we will see, the list is great at giving us candidates for the worst trains in the country, even if the train at its very bottom was a dud. And with those candidates it gives us, we can use the rest of our data to see if the trains are truly as bad as they seem, and if so, how Amtrak should be looking into fixing them.

Second, even if train 499 wasn't truly awful, what if Amtrak did want to devote some resources into improving it? What does typically cause it to be delayed? And to answer that, we can look at its delay metrics:

Delay Reason	Delay Minutes	Percentage
Initial Terminal Delay	762	30.346476
Miscellaneous Delays	348	13.859020
Unused Recovery Time	339	13.500597
Slow Order Delays	259	10.314616
Locomotive Failure	159	6.332139

Far and away, the most common cause for train 499's delays are Initial Terminal Delays – delays that are Amtrak's responsibility and that, according to the FRA's Service Quality Reports, are a "Delay at initial terminal due to late arriving inbound trains causing late release of equipment" [32]. This does not seem to be train 499's fault; thus, if Amtrak wants to improve train 499's performance, the best thing it can do is make sure that the trains arriving at the station before train 499 leaves, trains that train 499 needs equipment from, get to the station on time. Apart from that, the largest cause of delay for train 499 that isn't just "Miscellaneous Delays" is Unused Recovery Time, another type of delay that is Amtrak's responsibility. According to the FRA's Service Quality Reports, this is "Wait for departure time," or time that the train is stuck waiting at a station until it can depart [32]. Thus, to improve performance Amtrak should make sure that their trains are able to depart as soon as they are ready, whether that be through making sure staff is prepared sooner, getting all the customers properly on board quickly, or improving whatever other reasons may be causing this delay. Finally, train 499 is a rare train on tracks that are not owned by Amtrak that does not suffer from a high percentage of freight train interference; the percentage sits at only 5.77% despite the majority of its tracks being owned by the Massachusetts Department of Transportation. This is good for train 499, as it clearly rarely has to worry about freight trains. However, there would be no way for Amtrak to carry this good fortune to other parts of the country as the Massachusetts Department of Transportation clearly only manages tracks in Massachusetts, and, being a government entity, they probably work to prioritize passenger trains over freight trains as is legally required.

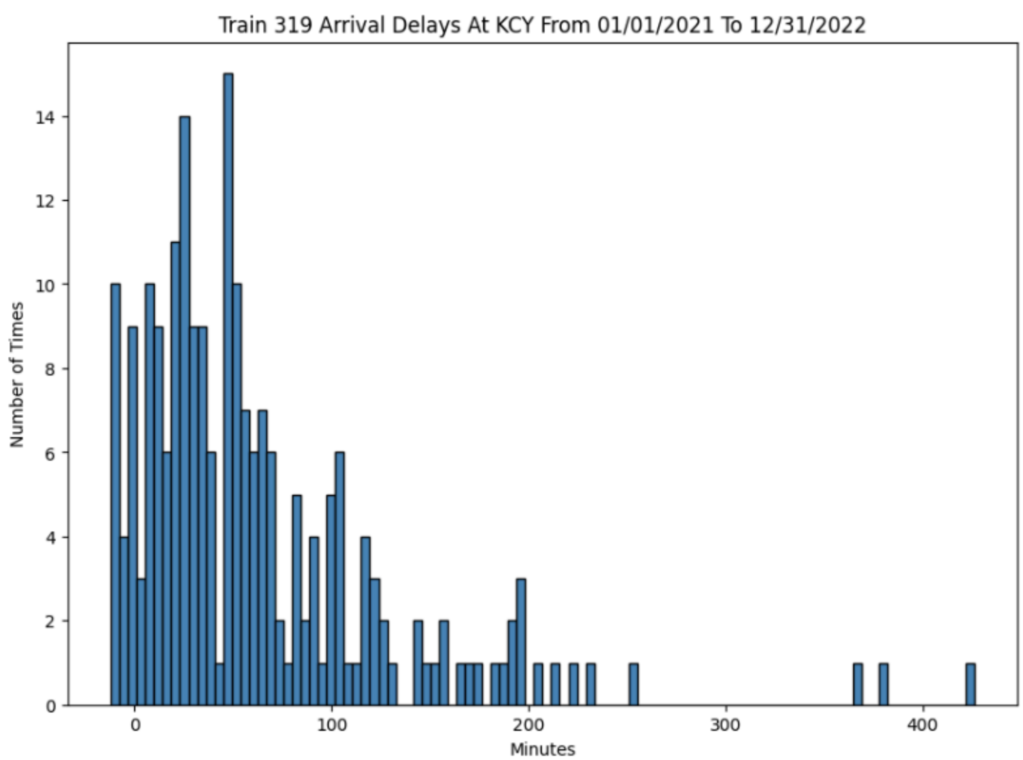
Host Railroad Name	Delay Reason	Delay Responsibility	Delay Minutes	Host Railroad Train Miles
Amtrak	Freight Train Interference	Host Railroad	48	32.4
Massachusetts DOT	Freight Train Interference	Host Railroad	97	3809.2

The Massachusetts DOT caused very little freight train delay for train 499 between July 2021 and June 2022, despite the significant mileage that the train traveled on their tracks.

4.1.2: Train 319

The second to worst train on our list of trains by median delay per mile from July through June is train 319. Train 319 starts its route by going from Chicago to St. Louis as part of Amtrak's Lincoln Service, before going from St. Louis to Kansas City as part of Amtrak's Missouri River Runner Service; this makes it a rare Amtrak train that runs as a part of multiple services (though metrics for the two services are occasionally lumped together as one), as well as one of two trains that are a part of the combined Lincoln/Missouri service (the other being 318 in the opposite direction). Thus, 319 appears to be the worst long-distance train on our list, but as we just saw with train 499, appearances can be deceiving. Is train 319 truly as bad as our list makes it out to be?

Yes, it is.




```
Mean 62.30373831775701
Median 46.5
Variance 4523.660079482925
```

Train 319, in all of 2021 and 2022, was an average of 62 minutes late to its final stop, and its median delay was not much better at 46.5 minutes. This helps to keep it very low in our list of trains by median delay per mile for all of 2021 and 2022 as well, where it sits at sixth from the bottom. Further, the customer on-time performance for train 319 is predictably not great either, sitting at an abysmal 24.32% for the third fiscal quarter of 2022.

Train Number	C OTP
319	0.243247

Even this data is deceiving though, as train 319, as well as its opposite train 318, didn't actually begin running until May 23, 2022; thus, it has not only had its current poor performance come in an incredible short time frame, but an incredibly recent time frame as well; in other words, train 319 is a very new train, and any ways we can find to improve it could be very useful for its future performance. However, this does also cause a bit of an issue we were trying to avoid, as with the train starting in late May 2022, we only have delay metrics for its runs from its start through June 2022, and what caused the train to be late in its first month of running may not be consistent for what continues to keep it delayed now, almost one year later.

That being said, we can still see what the primary causes of delay were for train 319, from its start on May 23rd until the date of the end of our delay metric data on June 30th, and perhaps gain some insight into what could continue to delay it today.

Delay Reason	Delay Minutes	Percentage
Freight Train Interference	1433	26.346755
Passenger Train Interference	596	10.957897
Slow Order Delays	595	10.939511
Passenger Related - Disabilities	467	8.586137
Passenger Related	381	7.004964
Weather-Related	378	6.949807
Signal Delays	324	5.956977

Here, we begin to see the reason for delay that often is the plight of long-distance trains: freight train interference. This was the biggest cause of delay for the budding train 319, and unfortunately there is little Amtrak can do about it as they do not own the tracks in the Midwest where train 319 operates. Its other top two causes, passenger train interference and slow order delays, unfortunately also fall as the responsibility of the host railroad. Thus, Amtrak themselves would not be able to fix most of train 319's delays alone, and the responsibility for improving the train's performance in these areas would fall on the host railroads, and as is seen below, primarily Union Pacific.

Host Railroad Name	Delay Reason	Delay Responsibility	Delay Minutes	Host Railroad Train Miles
Kansas City Terminal	Freight Train Interference	Host Railroad	74	231.8
Terminal Railroad Assn. Of St. Louis (TRRA)	Freight Train Interference	Host Railroad	47	251.7
CN - IC (Former GTW and IC)	Freight Train Interference	Host Railroad	120	1357.2
Union Pacific	Freight Train Interference	Host Railroad	1192	19735.7

However, there is one major area that Amtrak has control over where train 319 has been consistently delayed: passenger delays. In its first month of operation, about 15% of train 319's delay was caused by some form of passenger delay, whether it came from normal passenger operations or operations related to disabled passengers. Thus, having the train improve in its handling of passengers – whether it be by making sure everyone gets on the train in a timely manner, ensuring that baggage gets where it needs to be, or providing support and accessibility for disabled passengers – would definitely help to improve the performance of the train.

However, regardless of whether these issues have been addressed or not, in the time from our delay metrics to now, train 319 does seem to have improved significantly; according to ASMAD, from September 2022 through February 2023, compared to the old mean and median arrival delay of 62 minutes and 47 minutes respectively, the train now has a mean and median arrival time of 38 minutes and 26 minutes respectively. This improvement is backed up by our hypothesis tests as well, as seen below, which gives a difference in average delay this big essentially a 0% chance of happening without a significant change in the train's expected performance.

```
Train #: Train 319
Destination: KCY
Time Frame 1: 01/01/2021 - 08/31/2022
Time Frame #1 Average Delay: 88.66326530612245 Minutes
Time Frame 2: 09/01/2022 - 03/31/2023
Time Frame #2 Average Delay: 39.3502538071066 Minutes
Test Statistic: 5.487104027679816
P-value: 1.9777522561784762e-07
If the p-value is sufficiently small,
 time frame #1 is significantly more delayed than time frame #2 on average.
```

Perhaps Union Pacific adjusted through the growing pains of a new train so the train would be interfered less by freight and other trains in the future. Perhaps Amtrak adjusted by improving the handling of their passengers. Perhaps the train just needed a bit of time to start running smoothly. Either way, train 319's updated arrival delays show that while it's better now than it was, it's still not perfect, so perhaps as more delay metrics are released we can use them and our tools to find how train 319 can continue to improve.

4.1.3: Train 1014

Our next train, train 1014, is an Amtrak train that runs from Klamath Falls, Oregon to Seattle, Washington. The 3rd train from the bottom of the list of median normalized delays from July to June, a quick look at our list from 2021-2022 gives us an immediate surprise: the train is nowhere to be found. Immediately upon seeing this, my first thought was, “do we have another train 499, with a bad stretch at the wrong time?”

But then I did a little more digging. And not only is train 1014 not visible on our bigger list of trains; with a sample size of less than 104, it’s not on the list at all. And some more digging on ASMAD gives us why; this train only ran from July 16th, 2021 through August 23, 2021. It’s probably not worth it to analyze a train that ran for such a short period of time, but a question stuck in my head regardless; why did it run for so short?

When I started searching for an answer, it took me a bit to find any leads. The best I could find was on the Amtrak Unlimited Discussion Forum from 2008 discussing 1014 running then, as a regional replacement for train 14 (the Amtrak train from Los Angeles to Seattle) due to Union Pacific needing to repair broken tracks; in this case train 14 was taking passengers from Los Angeles to Klamath Falls, Oregon, before said passengers would take a bus to Eugene, Oregon to avoid the broken tracks and finish the route on a temporary train 1014 traveling from Eugene to Seattle [6]. Perhaps something similar had happened in 2021, where damaged tracks had caused an inability for train 14 to finish its whole route, necessitating train 1014 to continue the route a bit later and finish the trip to Seattle.

This idea was enough for me to find the next piece of the puzzle: a Facebook video from July 2021 of train 1014 leaving one of its stops. While the video itself didn't actually help me, it was the video's description that gave me all I needed: "Fires have destroyed the UP line in Northern California and a bus bridge meets up with this train in K Falls" [38].

With this information, all I needed was a bit more research for everything to fall into place. In June and July 2021, wildfires in Northern California heavily damaged Union Pacific railroads, rendering trains unable to travel on them [10] [12] [33]. Thus, Amtrak had to make changes to its services – having train 14 run from Los Angeles to Sacramento, busing passengers from Sacramento to Klamath Falls, and setting up train 1014 for travel from Klamath Falls to Seattle – until the fires were under control and the tracks were repaired in August [7] [33] [34].

In this time that it ran, train 1014 did not perform all that well, accumulating an average arrival delay at Seattle of 56 minutes, with a median of 48 minutes. Further, to no one's surprise, much of train 1014's reasons for its delay in its short running time fell exactly in line with the circumstances that led to its creation.

Delay Reason	Delay Minutes	Percentage
Freight Train Interference	959	19.941776
Slow Order Delays	752	15.637347
Weather-Related	539	11.208151
Passenger Train Interference	444	9.232689
Police-Related	422	8.775213
Routing	399	8.296943

In addition to the typical freight train interference experienced by trains outside of Amtrak's track ownership, train 1014 fell victim to higher than usual weather-related, police-related, and routing delays, which were almost certainly related to the fires that were going on around where it was operating. This is further supported by train 14, which usually covers the track from Klamath Falls to Seattle as part of its whole route from Los Angeles to Seattle, having very little of these types of delay.

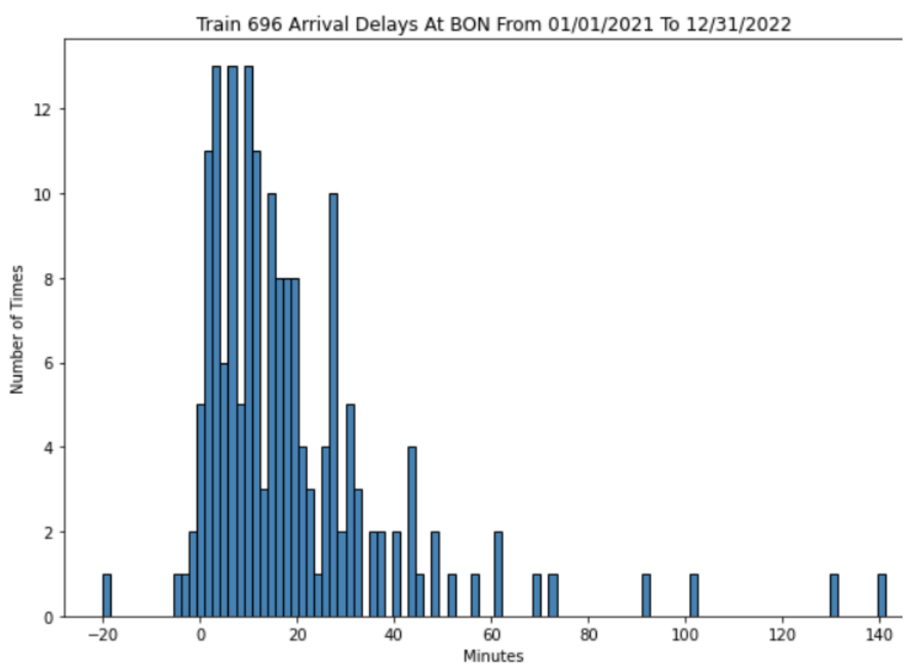
Police-Related	3373	3.320961
Weather-Related	2189	2.155228
Routing	2166	2.132582

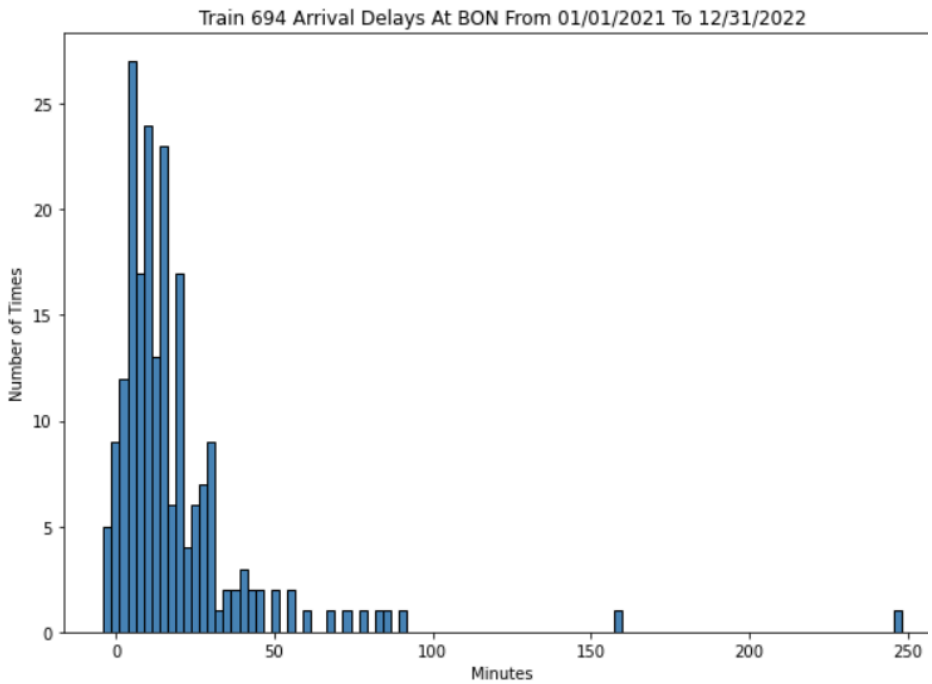
Delay minutes and percentages for the applicable delay reasons for train 14.

Thus, we can see that the unique causes of delay for train 1014 were caused by the circumstances of its creation. Therefore, even if Amtrak had prior knowledge that these causes would delay train 1014, and if any similar situation if the future were to require a revival of the train, it's unlikely that Amtrak would be able to substantially improve the train's performance considering the precarious circumstances that would necessitate the train's existence.

4.1.4: Train 696 and Train 694

Train 696 and train 694, the 4th and 5th trains from the bottom of our list, are both trains that cover the same route; as part of Amtrak's Downeaster service, these trains travel from their origin at Brunswick, Maine to their destination about 120 miles away at Boston's North Station. These two trains also consistently appear near the bottom of our lists for the most delayed trains; train 696 appears at 3rd from the bottom of our 2021-2022 list of median train delay per mile and 4th from the bottom in the July 2021 through June 2022 list of the same metric, while train 694, through performing slightly better, appears at 7th from the bottom on the former list and 5th from the bottom on the latter. And while both of these trains are not quite as low on the mean train delay per mile lists, they are still both bottom 15 in both time frames for these lists. Looking at the distribution of these trains' delays in 2021 and 2022 also appears to back up their placements in these lists: they are clearly late a lot of the time.





Further, perhaps expectedly due to their similar routes and placements in my lists, neither of these trains were more significantly delayed than the other in 2021-2022, as is shown by my hypothesis testing with a p-value of 0.874. Given their similarities, it will thus be useful to analyze both of them and see if we can find a way to improve both of these highly delayed trains at the same time.

```
Train #1: Train 696
Destination #1: BON
Train #1 Average Delay: 18.971264367816094 Minutes
Train #2: Train 694
Destination #2: BON
Train #2 Average Delay: 18.59313725490196 Minutes
Time Frame: 01/01/2021 - 12/31/2022
Test Statistic: 0.15885205292320984
P-value: 0.8738707249634674
If the p-value is sufficiently small,
 train #1 is significantly more delayed than train #2 on average.
```

However, there is an interesting aspect of these trains being often late; they are a part of Amtrak's Downeaster service, which by Amtrak's own delay per 10K miles metric is one of their better services. In fact, out of Amtrak's 34 services they have delay per 10K miles metrics for, the Downeaster is the 11th best service by delay per 10K miles, which is certainly not the placement you would expect to see from a service that appears to have two of Amtrak's worst trains.

So are train 696 and train 694 trains that are just worse than the rest of the Downeaster service, or has something gone wrong in our collection and analysis of the data? A good way to answer this question is to look at the customer on-time performance of trains 696 and 694 and compare them to the customer on-time performance of the whole Downeaster service; if these trains are the worst that a generally good service has to offer, as they appear to be, then their customer on-time performance should be much worse than the service as a whole.

Service	C OTP
Downeaster	0.846586

Train Number	C OTP
696	0.590683

Train Number	C OTP
694	0.63951

As expected, this is in fact the case! The Downeaster has a very respectable customer on-time performance percentage of 84.66%, which is comparable to train 499 as we saw earlier. Meanwhile, train 696 and train 694 have much worse performance, with percentages of 59.07% and 63.95% respectively. Sure, these aren't the worst performance statistics possible, but they definitely aren't what I'd consider good. Why, then, are these trains so much worse than the rest of their service, and how can they be brought up to par with the rest of the Downeaster?

Perhaps we can find an answer in the delay metrics of these trains versus the delay metrics of the Downeaster as a whole. There may be something more likely to delay trains 696 and 694 than the rest of the trains on the Downeaster, and if we're lucky it could be something that Amtrak can control and fix. Below we can see the most common causes of delay for the whole Downeaster service from July 2021 to June 2022 on top, with the delay causes for trains 696 and 694 below that, with train 696's delay reasons on the left and train 694's delay reasons on the right.

Downeaster Delay Reasons

Delay Reason	Delay Minutes	Percentage
Slow Order Delays	13453	18.606935
Passenger Train Interference	12287	16.994232
Unused Recovery Time	10870	15.034370
Signal Delays	10063	13.918203
Commuter Train Interference	5935	8.208738
Freight Train Interference	5553	7.680392

Train 696 Delay Reasons

Delay Reason	Delay Minutes	Percentage
Passenger Train Interference	583	18.940871
Commuter Train Interference	454	14.749838
Signal Delays	427	13.872645
Slow Order Delays	408	13.255361
Freight Train Interference	378	12.280702
Unused Recovery Time	152	4.938272

Train 694 Delay Reasons

Delay Reason	Delay Minutes	Percentage
Passenger Train Interference	733	27.401869
Slow Order Delays	443	16.560748
Signal Delays	424	15.850467
Commuter Train Interference	237	8.859813
Freight Train Interference	193	7.214953
Passenger Related	133	4.971963

The biggest difference between the Downeaster as a service and trains 696 and 694 specifically is the impact of train interference. About 46% of train 696's delay is caused by some kind of train delay, as well as about 43.5% of train 694's delay, while only about 33% of the whole service's delay is caused by such factors. Passenger train interference clearly plays a larger role in the two delayed trains as well, appearing more often in both trains than overall in the Downeaster. So clearly, these trains seem to be more delayed because they run into more trains than the rest of their service.

However, this doesn't seem to paint the whole picture. Why do these two trains run into more track usage than the rest of the service? Is there anything else that's different about train 696 and train 694 compared to the rest of the Downeaster that may cause these differences in delay? And after a good bit of searching, I was able to find two major differences between these two disrupted trains and the rest of their service that could easily contribute to their lateness. Firstly, train 696 and train 694 are weekend trains. The Downeaster as a service runs daily, regardless of the day of the week, and weekday trains such as those in the 680s tend to run quite well, as opposed to train 696 and 694. In addition, trains 696 and 694 operate around midday, while early and late day trains such as train 690 and train 698 tend to run better than them. Why is it that the trains that run during the week or early or late in the day tend to run better than trains 696 and 694? Perhaps the combination of train 696 and 694 being weekend trains that operate around midday contribute to there being more traffic on the tracks when they look to run, both from other passenger trains filled with people traveling on the weekend and from freight trains running at their peak hours. When people are off from work for the weekend, they may look to take a trip, and traveling on a local passenger train seems like an excellent way to do this. In addition, many freight trains probably do not stop operating on the weekend, and they'll be primarily operating in the daylight if they are able. Thus, it seems to me that train 696 and train 694 are as delayed as they are because they run when the tracks they use are at their busiest, in the middle of the day on the weekend, and this causes these trains to run into more interference from other trains and thus be more delayed than the rest of the Downeaster. If this is correct, then there wouldn't be much for Amtrak to do to improve these trains' performance; if you're running trains at peak hours for track business, you're going to run into more delays, which seems to be exactly what is happening to these trains.

4.2: Personally Selected Trains for Analysis

4.2.1: Trains 1 through 6

Contrary to the trains we analyzed so far, none of trains 1 through 6 fall near the bottom of any of our lists of trains by delay per mile. Instead, these six trains are almost always sitting at or near the very bottom of our lists of trains by net delay, indicating that in terms of average and median delay at their destinations, these six trains are the worst trains in the country.

Thus, it shouldn't be surprising to learn that trains 1 through 6 compose three of Amtrak's long-distance train lines, with trains 1 and 2 connecting Los Angeles and New Orleans, trains 3 and 4 connecting Los Angeles and Chicago, and trains 5 and 6 connecting Emeryville, California and Chicago. With the vast distances these routes have to travel, there is plenty of time for them to get hit with lots of delay. And for all six of these trains, if they're going to get hit by any type of delay, it's probably going to be caused by the host railroads, and it's most likely going to be freight train interference.

Train 1 Delay Responsibilities and Reasons

Delay Responsibility	Delay Minutes	Percentage
Host Railroad	63570	71.058897
Amtrak	18581	20.769944
Neither	7310	8.171158

Delay Reason	Delay Minutes	Percentage
Freight Train Interference	42083	47.040610
Slow Order Delays	7751	8.664111
Routing	5713	6.386023

Train 2 Delay Responsibilities and Reasons

Delay Responsibility	Delay Minutes	Percentage
Host Railroad	57496	69.607748
Amtrak	15865	19.207022
Neither	9239	11.185230

Delay Reason	Delay Minutes	Percentage
Freight Train Interference	34826	42.162228
Routing	7829	9.478208
Slow Order Delays	7827	9.475787

Train 3 Delay Responsibilities and Reasons

Delay Responsibility	Delay Minutes	Percentage
Host Railroad	71195	61.167768
Amtrak	28354	24.360572
Neither	16844	14.471661

Delay Reason	Delay Minutes	Percentage
Freight Train Interference	31385	26.964680
Slow Order Delays	13404	11.516156
Unused Recovery Time	10177	8.743653

Train 4 Delay Responsibilities and Reasons

Delay Responsibility	Delay Minutes	Percentage
Host Railroad	71395	65.444762
Amtrak	24671	22.614857
Neither	13026	11.940381

Delay Reason	Delay Minutes	Percentage
Freight Train Interference	32837	30.100282
Slow Order Delays	12952	11.872548
Signal Delays	10213	9.361823

Train 5 Delay Responsibilities and Reasons

Delay Responsibility	Delay Minutes	Percentage
Host Railroad	102872	65.687577
Amtrak	30006	19.159941
Neither	23730	15.152483

Delay Reason	Delay Minutes	Percentage
Freight Train Interference	49918	31.874489
Slow Order Delays	21605	13.795592
Passenger Train Interference	10338	6.601195

Train 6 Delay Responsibilities and Reasons

Delay Responsibility	Delay Minutes	Percentage
Host Railroad	103729	66.212818
Amtrak	31310	19.985957
Neither	21621	13.801226

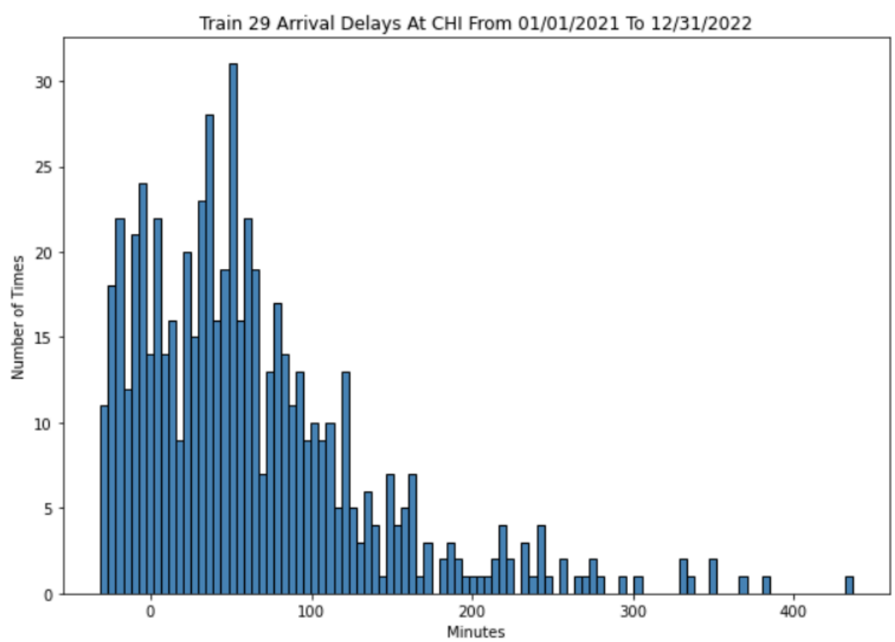
Delay Reason	Delay Minutes	Percentage
Freight Train Interference	49957	31.888804
Slow Order Delays	22410	14.304864
Signal Delays	9664	6.168773

There is not too much more to say about these trains; they are the epitome of long-distance trains getting delayed by freight and being hurt by running on tracks not owned by Amtrak. I do find it interesting, however, how none of them are near the bottom of the delay per mile lists; despite the intense delay they take on throughout their routes, they still run well enough on a per-mile basis to not perform as poorly as many other Amtrak trains.

4.2.2: Train 29 and Train 30

Train 29 and 30 are the trains of Amtrak's Capitol Limited service, a service that runs trains between Washington D.C. and Chicago, with intermediate stops in cities such as Pittsburgh and Cleveland. In my list of delayed trains, they initially caught my attention as having the largest sample size of any trains near the bottom of the list, a fact that makes sense considering the nearly daily runs that these trains take along their route; over 600 of these trains ran between 2021 and 2022, with over 300 running between July 2021 and June 2022. However, what initially started as curiosity – *how can trains that run this many times continue to run so poorly?* – quickly turned into horror as I began my investigation of the trains.

Perhaps the most damning statistic on this train is the first I looked at: train 29's distribution of arrival delays at its endpoint between 2021 and 2022:




```
Mean 62.770358306188925
Median 49.0
Variance 5361.189935702235
```

Train 29, over all of 2021 and 2022, had a median arrival delay of 49 minutes, and I can say that train 30, with a median delay of 40 minutes, was not much better. Further, perhaps as proof that both of these trains contain equal issues, hypothesis testing fails to find that these trains had a significant difference in their average delay between 2021 and 2022, with a p-value of 0.364. In other words, *both* of these trains perform *really* poorly.

```
Train #1: Train 29
Destination #1: CHI
Train #1 Average Delay: 62.770358306188925 Minutes
Train #2: Train 30
Destination #2: WAS
Train #2 Average Delay: 58.78896103896104 Minutes
Time Frame: 01/01/2021 - 12/31/2022
Test Statistic: 0.9079978635644432
P-value: 0.3640589002307515
If the p-value is sufficiently small,
 train #1 is significantly more delayed than train #2 on average.
```

The delays of these trains are abysmal, but it is important to remember that trains 29 and 30 are long trains, taking between 16 and 17 hours to get to their destination when they aren't delayed. This gives them plenty of opportunity to get more and more delayed as they continue throughout their whole journey. However, these trains do still appear low on our list of trains' delay per mile. And they don't only appear low on our list; the Capitol Limited is the second worst Amtrak service by Amtrak's official delay per 10K miles metric, and can be seen near the bottom of the list below.

Auto Train	128241	1232375.2	0.104060
Texas Eagle	194168	1791838.2	0.108362
Lincoln / Missouri	9754	86586.4	0.112650
Capitol Ltd	120981	1073830.4	0.112663
Sunset Ltd	155512	1199037.6	0.129697

And if you needed even more proof about the poor condition of these trains, the service's customer on-time performance has you covered there as well.

Service	C OTP
Capitol Ltd	0.28151

Ok, so it's clear that train 29 and 30 are bad. They're constantly delayed, among the worst trains in the country in terms of Amtrak's official delay per mile, and only 28% of their customers reach their destination on time. What causes the delay that makes these trains run so poorly?

Train 29 Delay Reasons

Delay Reason	Delay Minutes	Percentage
Freight Train Interference	35364	49.042422
Unused Recovery Time	6119	8.485765
Routing	5228	7.250135
Passenger Train Interference	5142	7.130871
Slow Order Delays	2634	3.652803

Train 30 Delay Reasons

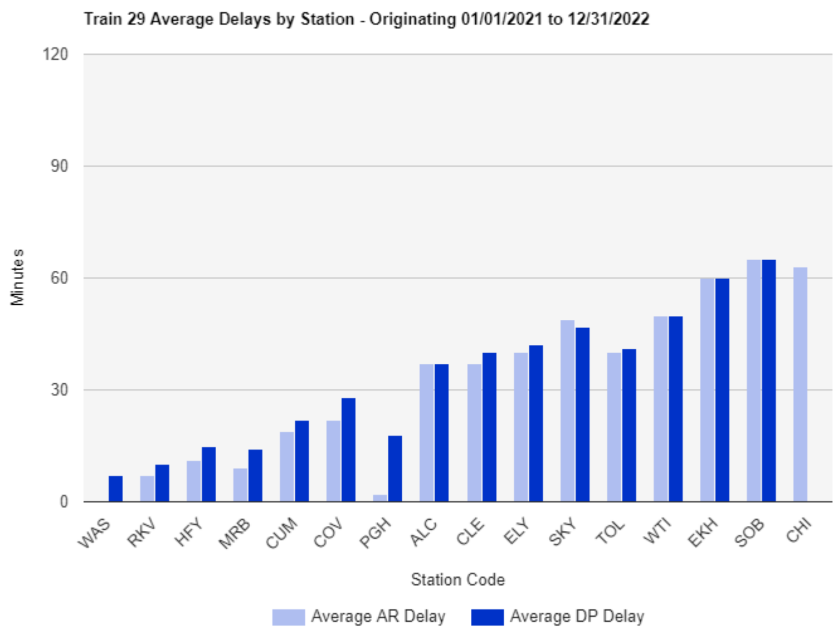
Delay Reason	Delay Minutes	Percentage
Freight Train Interference	29234	45.880285
Routing	8065	12.657334
Slow Order Delays	4129	6.480116
Crew & System	3161	4.960922
Passenger Train Interference	3091	4.851062

Far and away the greatest causes of delay for train 29 and train 30 is freight train interference. Wonderful. And we can even see which rails contribute the most to this issue for the service.

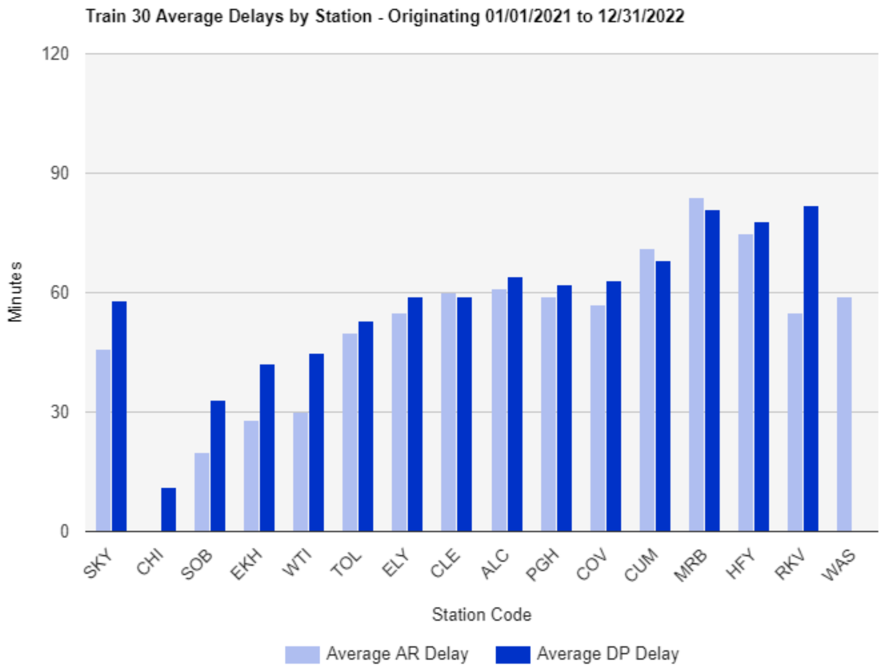
Host Railroad Name	Delay Responsibility	Delay Minutes	Host Railroad Train Miles	Percentage	Per 100 Miles
Norfolk Southern	Host Railroad	59464	330859.2	92.052386	17.972600
CSX Corporation	Host Railroad	5124	203372.8	7.932134	2.519511
Amtrak	Host Railroad	10	538.2	0.015480	1.858045

The vast majority of freight train delay on Amtrak's trains 29 and 30 comes on rails owned and operated by Norfolk Southern. In fact, despite having only about 1.6 times more railroad miles on the route of these trains than the CSX Corporation, Norfolk Southern contributes over 11 times the freight train delay.

Thus, it seems clear that the most common cause of delay for trains 29 and 30 is freight train delay caused by Norfolk Southern, which unfortunately Amtrak cannot fix. However, this isn't the only interesting part of these trains' delays. By looking at train 29's average delay at each station, we can notice something else very interesting.



Shockingly, by looking at its graph one can see that train 29 does average reaching one of its stations only 2 minutes late. And this isn't just at any random stop, this is at one of the route's most important: Pittsburgh. However, after reaching Pittsburgh, everything for the train falls apart. And this trend can be seen for train 30 too, just in the opposite direction.

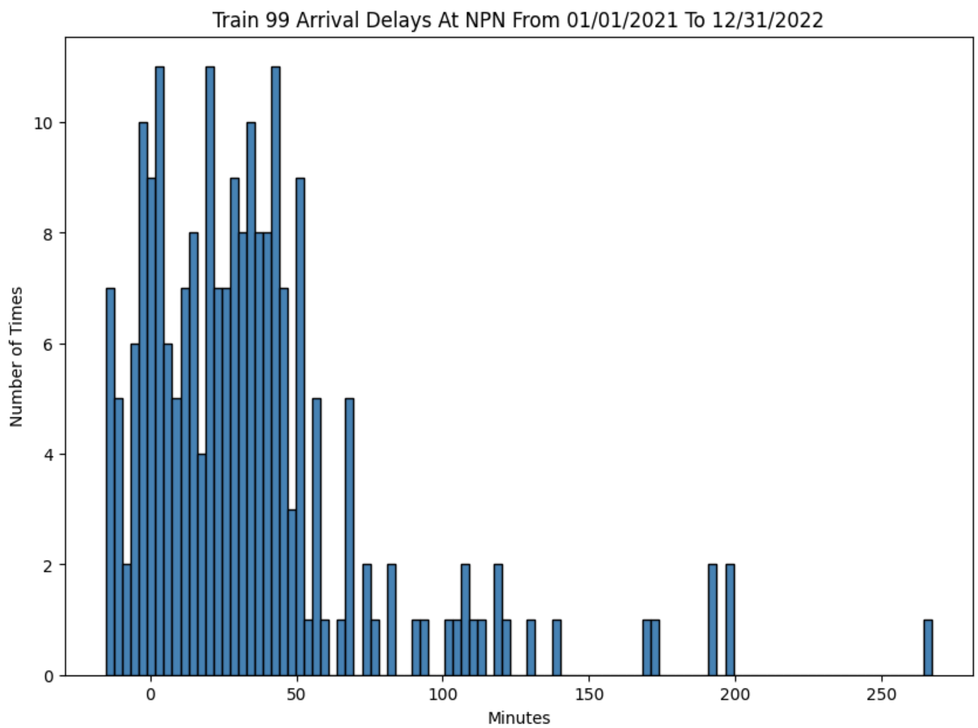


Ignoring the initial bar for SKY/Sandusky, Ohio (which should actually be between the bars for TOL/Toledo and ELY/Elyria, but is at the beginning instead due to an ASMAD error), one can see that train 30 gains most of its delay in the beginning of its route, as it travels from Chicago to Pittsburgh. And once it hits Pittsburgh, it too runs well along the stretch from Pittsburgh to D.C., and actually fails to gain (or lose) any delay on average between the two stops.

Thus, the truth of the Capitol Limited becomes clear; the train is killed by Norfolk Southern freight train delay between its stops of Pittsburgh and Chicago. And while this delay is not directly in Amtrak's control, if they want to fix the train connecting the nation's capital and the Windy City, they can at least see who they need to appeal to and where, in hopes that Norfolk Southern can give more priority to passenger trains over freight as they are legally required to.

4.2.3: Train 99

Train 99 is an Amtrak train that travels from Boston, Massachusetts, to Newport News, Virginia. Running primarily on the weekends, this train is scheduled to take just over 12 hours to complete its route from start to finish, though it usually takes a bit longer than scheduled, averaging a median delay of 30 minutes to its end in 2021 and 2022.




```
Mean 35.46296296296296
Median 29.5
Variance 1821.9523319615912
```

However, despite not being at the very bottom of any of our delayed train lists – train 99 falls at 13th from the bottom for our list of trains by delay per mile from July to June – what makes train 99 interesting is that it is a part of the Northeast Regional. Not only is the Northeast Regional one of Amtrak's best performing services, in 3rd place on our list of services by delay per 10K miles, but it also runs in the northeast where Amtrak owns the train tracks. That means, unlike with previous trains such as train 29 and 30, any delay that comes from the host railroad should be covered and fixable by Amtrak since they are the host railroad. And if Amtrak should be able to fix any causes of delay to train 99, then what causes it to still be so delayed?

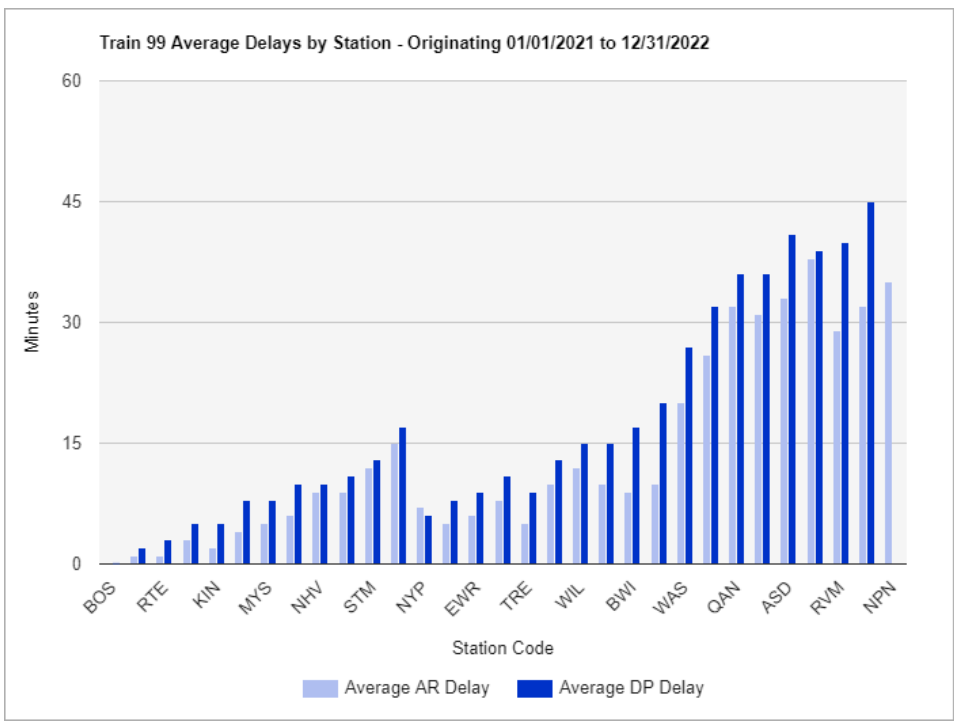
Delay Reason	Delay Minutes	Percentage
Weather-Related	1359	14.196177
Slow Order Delays	1194	12.472579
Passenger Related	833	8.701556
Routing	774	8.085240
Servicing	768	8.022563
Passenger Related - Disabilities	568	5.933354
Commuter Train Interference	539	5.630419
Freight Train Interference	531	5.546851

Train 99's most common reasons for delay are weather, a factor out of Amtrak's control, slow order delays, a host-railroad caused delay reason that Amtrak should be able to fix if train 99 is running on their tracks, and passenger related delays, which are Amtrak's responsibility regardless of whether or not they own the tracks. This seems good – Amtrak should be able to fix these problems – but it's still weird how train 99 performs so much worse than the rest of the northeast regional. Maybe looking at the reasons for delay for the whole Northeast Regional can give us some insight into why this one train is so much worse than the rest of the service.

Delay Reason	Delay Minutes	Percentage
Slow Order Delays	58030	12.782135
Weather-Related	44828	9.874161
Commuter Train Interference	43281	9.533407
Unused Recovery Time	35287	7.772587
Locomotive Failure	26429	5.821455
Freight Train Interference	24302	5.352946
Signal Delays	23144	5.097876

These delay reasons look very similar to train 99's, as just like before slow order delays and weather are the top two delay causes for the whole Northeast Regional. The whole region has less passenger-related delays than train 99 as well – perhaps these passenger delays are the reason train 99 is as delayed as it is? Yet why would one train on the Northeast Regional have such a higher proportion of passenger issues than the rest of the service, assuming they're taking the same passengers? In addition, the whole service also has far less Routing and Servicing delays than train 99, which would be an issue caused by the host-railroad. If all the Northeast Regional runs on the same Amtrak owned tracks, why should this one train have so much more of this host-railroad caused delay?

Perhaps one more source of data can give us the clues to find what is different about this train: the delay for train 99 at every station it stops at.



Unexpectedly, Train 99 actually runs quite well in the beginning parts of its route; it averages an arrival delay of 15 minutes or below at every single stop before

Washington D.C., which is the majority of its route. Only after hitting D.C. does the delay of train 99 dramatically increase, which is strange. Does something change once the train hits D.C. that would cause it to become more delayed?

Unfortunately, something does change, and over the course of the research I've done, I've found it to be about the most significant thing that could change: the ownership of the tracks. As of 2021, in the Newport News section of the Northeast Regional, the tracks are not just owned by Amtrak, but by CSX, Metro-North, and Norfolk Southern as well [2]. As it turns out, Amtrak does not own every part of track that its Northeast Regional trains run on, and entering an area where its host-railroad is no longer Amtrak is what causes train 99's delay to skyrocket. In fact, over its entire route, we can see that all of the freight train interference for train 99 is caused by CSX, probably in this section of its route.

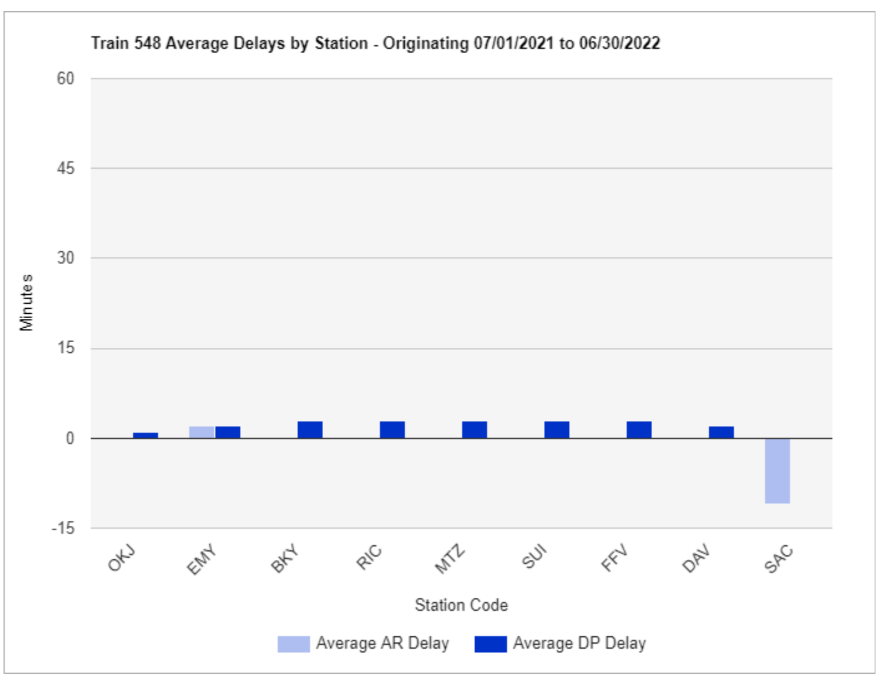
Host Railroad Name	Delay Reason	Delay Responsibility	Delay Minutes	Host Railroad Train Miles	Percentage
CSX Corporation	Freight Train Interference	Host Railroad	531	20088.0	100.0

Perhaps things could be changing for train 99 soon, however, as Virginia recently finalized a deal to buy over 200 miles of track from CSX for Amtrak [31]. In fact, as of 2019, Amtrak actually owned no tracks in the Newport News section of the Northeast Regional, and as we've already seen, that changed at least a little by 2021 [2] [36]. In addition, Amtrak can also still work to fix the passenger-related issues that are unique to train 99 compared to the rest of the Northeast Region, and that the host-railroad, whether it be Amtrak or another, should not have caused. Thus, even though the majority of train 99's delay currently comes from a region where Amtrak has little to no impact on the host-railroad caused reasons that delay it, there is a chance that this could change in the future and that train 99's performance could dramatically improve.

4.2.4: Train 548 and Other Early Trains

Finally, we have train 548. Train 548 is different from the trains we have seen so far; in all our delay-per-mile metrics, train 548 makes a surprising appearance. Instead of appearing at the bottom of the list, it appears at the top of every single one. That's right – according to our list, train 548 appears to be by far the best Amtrak train. Why is this?

It may seem that it would be difficult to explain why a train doesn't get delayed. Perhaps you can look for common delay reasons that do not apply to your given early train. And this could seem to be the case for train 548, which travels along its route in California from Oakland to Sacramento; it encounters very little freight train interference, only being delayed by it for 15 minutes between July 2021 and June 2022. However, diving a little deeper reveals something else. How is train 548 so early at its destination, if it has to leave on-time from its other stops? Train 548 has seven stops along its route, so how does it gain so much time? And when does it actually tend to depart from its other stops? For better or for worse, the ASMAD graph below gives us our answer.

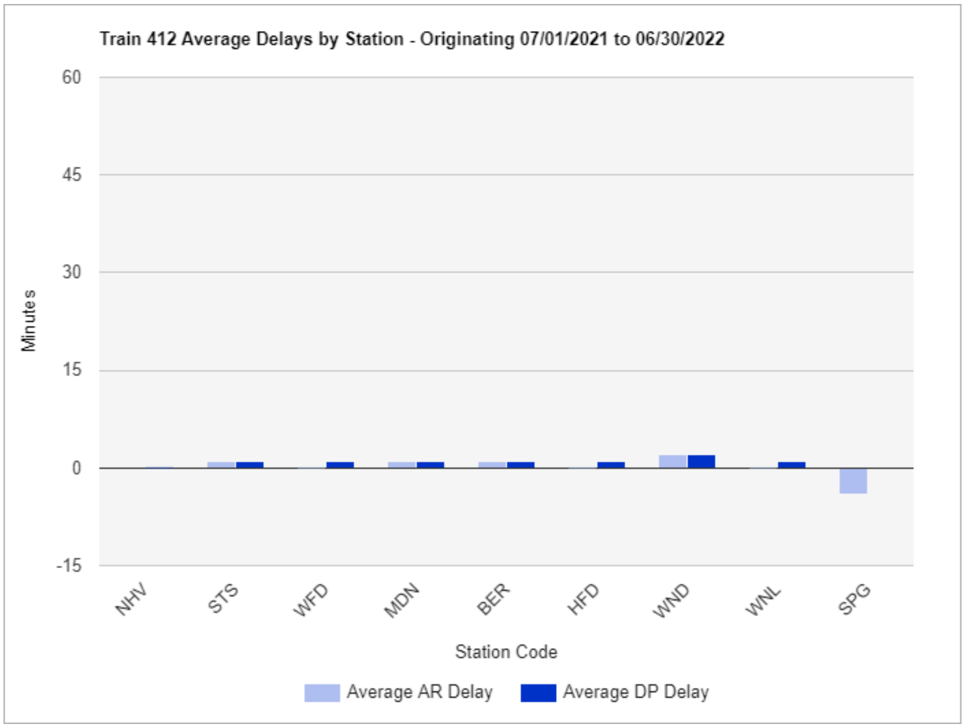


From July 2021 to June 2022, despite leaving an average of 2 minutes late from its second-to-last stop, train 548 arrived at its destination an average of 11 minutes early! Thus, the secret to being an early train is revealed; be a good running train with a really forgiving schedule. Don’t get me wrong, train 548 makes excellent time over the course of its entire run; averaging a departure delay of at most 3 minutes at any station over the course of a 70 mile journey is definitely impressive. However, if a train has a realistic schedule at its destination, it should not be able to go an average of 13 minutes faster than scheduled between its second-to-last and last stop. Realistically, train 548 should be scheduled to arrive at its final destination of Sacramento sooner than it currently is, and it's that scheduling that makes it seem like it runs better than every other train.

Unfortunately, this isn’t an issue that is just confined to train 548 either; of our 14 least delayed trains in our list of median delay per mile from July to June, 11 of them go from Oakland to Sacramento at some point in their route while also gaining massive time between their second to last stop and Sacramento. And of the three trains that are not a part of this, two of them are seasonal trains that travel from Denver, Colorado to the nearby Winter Park - Ski Resort with no intermediate stops, while the 3rd, train 1564 from Los Angeles to San Diego, departs an average of 6 minutes late from its third to last stop (and an average of 25 minutes late from its second to last stop, but with only 19% of the data available) to arrive at its destination an average of 4 minutes early. This means that to be at the top of our list of least delayed trains, it seems you have one option: gain a lot of time between your second to last and last stop so that it appears you arrived early.

But once you get past these first 14 trains, it’s trains number 15 and 16 on the list that finally seem to have earned their spots: train 412 and train 330.

Train 412 is a train that travels from New Haven, Connecticut to Springfield, Massachusetts, and between July 2021 and June 2022, it averaged arriving to and leaving from all nine stops on its route no later than 2 minutes late, reaching Springfield an average of 4 minutes early.



During this time the greatest factors to its delay were weather and cab car failures, though the fact that it's a weekday train that runs past 10PM probably also helped to negate traffic on the tracks that may have otherwise slowed it down.

Delay Reason	Delay Minutes	Percentage
Weather-Related	81	34.033613
Cab Car Failure	58	24.369748
Signal Delays	18	7.563025

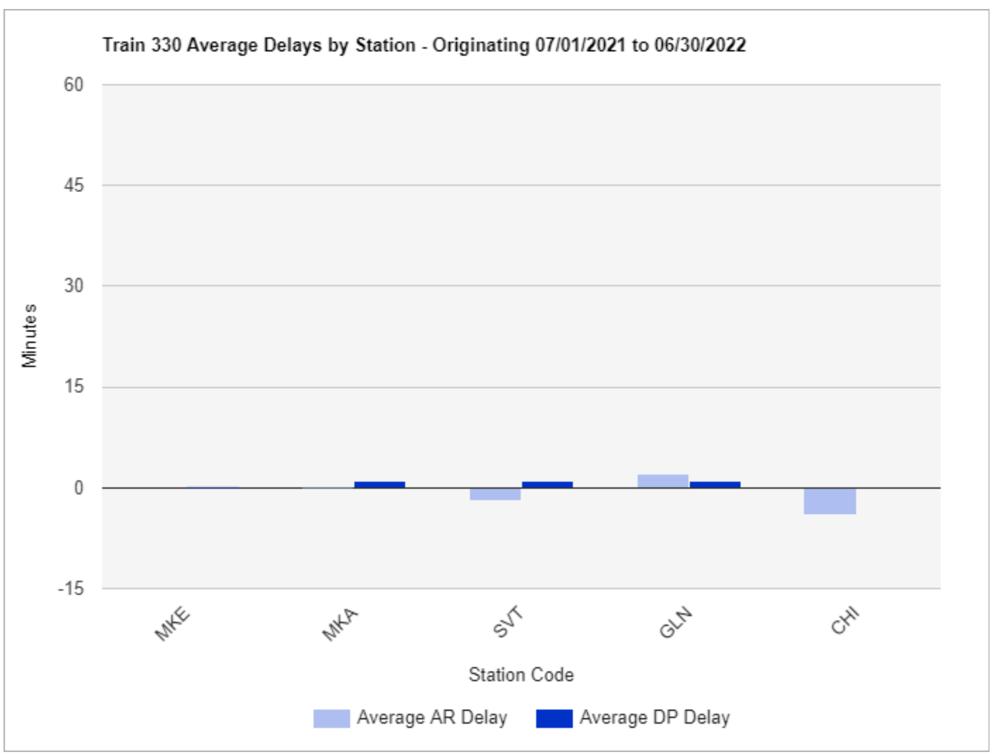
Train 412's top delay reasons from July 2021 to June 2022.

Incredibly, the delay metrics have train 412 as only accumulating 238 minutes of delay from July to June, as is seen below, despite the train running 135 times.

Delay Responsibility	Delay Minutes	Percentage
Amtrak	104	43.697479
Neither	85	35.714286
Host Railroad	49	20.588235

Delay minutes and responsibilities from July 2021 to June 2022 for train 412.

In addition to train 412, we also have train 330, an early morning train from Milwaukee to Chicago that runs before 8AM. With only five stops on its route, it did not average leaving any of its stops more than a single minute late, before getting to Chicago an average of 4 minutes early.



The biggest causes of delay to train 330 were unused recovery time and commuter train interference, which together contributed for over 75% of its delay minutes.

Delay Reason	Delay Minutes	Percentage
Unused Recovery Time	1387	42.030303
Commuter Train Interference	1101	33.363636
Signal Delays	228	6.909091

Train 330's top delay reasons from July 2021 to June 2022.

With this, it's easy to make a case for train 412 and train 330 as being the best run trains on all of Amtrak. And while both of them do have their advantages of when in the day they run, they are much more consistent throughout their routes and throughout their runs than any other train that could be in contention for such a title.

5. Conclusion

5.1: Results

In this thesis, we have accomplished three major goals. First, we identified the most delayed Amtrak trains at their final station from 2021 through 2022 as well as from July 2021 through June 2022, on the basis of mean net delay, median net delay, mean delay per approximate distance traveled, and median delay per approximate distance traveled. Second, we created tools to allow for easy analysis of these trains, allowing us to see why these trains tend to get delayed, if their delay is or is not possible for Amtrak to fix, how their delay compares to other similar trains, and if the trains actually do perform as badly as their delay statistics indicate. And third, we tested these tools on the worst performing trains we found from July 2021 through June 2022 by the median delay per distance traveled metric, thus discovering if and why these trains performed as badly as they seemed to while also verifying the ability of the tools we created to succeed in their purpose to be used as tools for Amtrak train and train delay analysis.

Furthermore, in our analysis of the most delayed trains, we did find that most of the causes for the delay of Amtrak's worst trains are causes outside of Amtrak's control. Thinking about it after the fact, this does make sense – if Amtrak could fix the problems with their most delayed trains, they probably would, making them no longer the most delayed trains. Further, as stated by Amtrak, freight train interference was the most common delay reason for the worst performing trains, but other host-railroad delay reasons such as slow order delays and other train interference were common as well. Meanwhile, one reason we found that caused the most delayed individual train runs, the weather, was not common in our reasons for typical train delay, though as seen by our analysis of train 99 it also was not completely absent. Finally, there were also places where Amtrak could improve the delay of select trains, such as by improving significant passenger-related delays; these reasons, however, were decently uncommon. Therefore, our analysis has shown Amtrak does a good job at minimizing the delay reasons that they would be responsible for impacting their trains, and for most if not all of Amtrak's most delayed trains it would be very difficult for them to make changes to significantly improve the trains' performance.

Thus, we have achieved all of our goals for this thesis in identifying Amtrak's most delayed trains, creating tools to help analyze Amtrak train delay, and verifying the usefulness of these tools by analyzing some poor performing Amtrak trains. Hopefully, these tools can persist to be used to analyze Amtrak trains and their delays, to continue to find what causes specific trains to be delayed and to find ways that their performance can be improved.

5.2: Areas for Further Research

There are multiple ways the research and tools presented above could be both built upon and improved upon, should one wish to do so.

The most obvious way further research could be pursued from this thesis is by using the tools offered by this thesis to analyze more trains. Only a small subset of the Amtrak trains that ran within our relevant time frames were actually analyzed within this thesis, and thus one could easily use the lists made of the most delayed trains, as well as the tools created to analyze them, to examine more trains that were not looked at in this report. This is further made possible since I have made the code used to create these lists and tools publicly available, with a link to the code being provided in the Appendix of this report. Perhaps one could analyze more very delayed trains to see exactly what causes their plights, or perhaps one could analyze trains more towards the middle of the list in hopes of finding trains with more Amtrak-caused and Amtrak-fixable delay.

Furthermore, one could make modifications to my code and methodology in order to change how the trains are ordered and analyzed. The most obvious way one could improve the lists I made is by obtaining the exact rail distance traveled by every train and applying that to the delay per distance traveled metric, instead of using an approximated distance as I did. While such an effort would definitely be time consuming, the results could be well worth it in making the lists of delayed trains more accurate (though as specified earlier in the report, to make it accurate the exact distance traveled would have to be found for every single train, as to avoid mixing trains with delay per distance travel statistics that were found in different ways).

In addition to looking at exact distance, one could also decide to analyze trains based on their total delay minutes accumulated, as opposed to analyzing them based on their delay at their final station as I did. I chose to analyze trains based on their delay at their final station in order to account for all the time they would both gain and lose throughout their whole route, as it felt more natural to base the performance of the trains off of how long it took them to complete their whole journey. However, one could prefer to base train performance on the total number of minutes they are delayed, regardless of if they make up that time later in their routes. In fact, the Delay Metrics from the FRA's Intercity Passenger Rail Service Quality and Performance Reports should contain all the data needed to make such an endeavor a reality, at least in finding and listing the trains with the most delay minutes. Thus, while delay minutes was not the exact metric I chose to focus on in this report, further research could definitely use such a metric to analyze train performance in a slightly different way.

Finally, one could take the tools I have made and apply them to much more specific scenarios than I did. For example, while I primarily examined the most delayed trains across the whole United States, one could hone their analysis onto a more specific region, such as the Northeast, a singular Amtrak service, such as the Northeast Regional, or a singular Amtrak route, such as trains that go from Boston to Washington D.C. While I did create some tools to make analysis in this manner easier, I didn't use them in my analysis, and further analysis could make much greater use of them by looking at more specific areas of Amtrak.

These are the ideas I have come up with that one could use to build off my research and continue to analyze train delay. I look forward to seeing how the analysis of

Amtrak trains and their delay continues to grow into the future, regardless of whether it is based off what I have suggested here, built off of what I have done in this report in some way I have not thought of, or started anew from some other idea or point of view.

5.3: Epilogue

On the evening of March 16th, 2023, I was in a car heading through Pennsylvania. Traveling on a road trip through the state with a couple of friends, we were currently driving through Erie, eager to make it to the beach of Lake Erie so we could enjoy the view of the sunset along the shores. And as we drove over some train tracks on our way, I couldn't help but think back to when I had been on those same tracks a year before, as a passenger on a train instead of in a car, as my trip from Princeton to California had begun.

I certainly knew a lot more about that trip now than I had then, and not only about the trains I had traveled on during it – though knowing that both legs of my trip were faster than average was pretty nice. I now also had way more information about Amtrak passenger trains than I ever thought possible. How they performed, the stories behind them, and the ways they could continue to improve; the possibility to see how these trains that I had dedicated so much time to would continue to evolve as time went on excited me to no end.

Yet as we arrived at the beach, I decided to put these thoughts in the back of my head. Train 5, 49, 319, 1014; these interests could wait. For now it was time to relax and enjoy what I had come here for; a chill time with some beautiful views and some great friends. I had come a long way to get to where I was now, and I knew that in the days ahead I would still have a long way to go. *But I am on my way*, I thought, as I stared into the horizon, always ready for the next adventure.

6. Appendix

All the code written for this report is publicly available on my GitHub at:

https://github.com/Aidan-Lynott/Aidan-Lynott-Senior-Thesis

Due to the vast amount of code written for this thesis, none of it is presented in this paper itself, except for its output. Thus, if you are interested in either viewing or using the code I wrote to create the lists of delayed trains or to analyze the most delayed trains, please use the GitHub link above.

7. Bibliography

[1] "Amtrak Fact Sheet: California Zephyr Service." *Rail Passengers Association*, Rail Passengers Association, https://www.railpassengers.org/site/assets/files/3440/27.pdf.

[2] "Amtrak FY21 Fact Sheet - State of Virginia." *Amtrak*, Amtrak Government Affairs, Aug. 2022, https://www.amtrak.com/content/dam/projects/dotcom/english/public/documents/corporate/statefactsheets/VIRGINIA21.pdf.

[3] "Amtrak Host Railroad Report Card 2021." *Amtrak*, National Railroad Passenger Corporation, https://www.amtrak.com/content/dam/projects/dotcom/english/public/documents/corporate/HostRailroadReports/Amtrak-2021-Host-Railroad-Report-Card.pdf.

[4] "Amtrak On-Time Performance Trends and Hours of Delay by Cause." *Bureau of Transportation Statistics*, U.S. Department of Transportation, https://www.bts.gov/content/amtrak-time-performance-trends-and-hours-delay-cause.

[5] Amtrak. "@COWinePress Our Train 5 (26jun) Was Ordered to Reversed Back to OTM Due to Flooded Tracks in That Area. Now Waiting for an OK to Proceed." *Twitter*, 27 June 2015, https://twitter.com/Amtrak/status/614941455761735680.

[6] Amtrakwolverine. "Service Alert: Coast Starlight Trains 14/1014 and 11/1011 -

Service Re." *Amtrak Unlimited*, Xenforo, 21 Apr. 2008, https://www.amtraktrains.com/threads/service-alert-coast-starlight-trains-14-1014-and-11-1011-%E2%80%94-service-re.22552/.

[7] "Bus Bridge Operating Overnight For Coast Starlight Passengers." *Amtrak in the Heartland*, WordPress, 21 July 2021, https://csanders429.wordpress.com/2021/07/21/bus-bridge-operating-overnight-for-coast-starlight-passengers/.

[8] "Developer Reference - Postal Code Radius & Distance Calculation." *IGEOCODE*, IGEOCODE, https://igeocode.com/developer/reference/great-circle-distance.

[9] "Flagstaff, AZ Weather History." *Weather Underground*, TWC Product and Technology LLC, https://www.wunderground.com/history/weekly/us/az/joseph-city/KFLG/date/2021-1-24.

[10] Franz, Justin. "UP Bridge Damaged by Fire Will Take Months to Rebuild." *Railfan & Railroad Magazine*, White River Productions, 16 July 2021, https://railfan.com/up-bridge-damaged-by-fire-will-take-months-to-rebuild/.

[11] "Freight Delays and Your Amtrak Service." *Amtrak*, National Railroad Passenger Corporation, https://www.amtrak.com/on-time-performance.

[12] Gailey, Brian. "Union Pacific Tracks Heavily Damaged in Lava Fire." *Klamath Falls News*, 8 July 2021, https://www.klamathfallsnews.org/news/union-pacific-tracks-heavily-damaged-in-lava-fire.

[13] "Illinois River Nearing Record Levels in Havana, Ill." *The Visual Journal*, The State Journal-Register, 2016, http://visuals.sj-r.com/illinois-river-flooding-at-havana-june-25-2015.

[14] "Intercity Passenger Rail Service Quality and Performance Reports." *Federal Railroad Administration*, U.S. Department Of Transportation, 28 Feb. 2023, https://railroads.dot.gov/rail-network-development/passenger-rail/amtrak/intercity-passenger-rail-service-quality-and.

[15] Juckins, Christopher. "Amtrak Stations List." *Amtrak Status Maps Archive Database*, https://juckins.net/amtrak_status/archive/html/stations.php.

[16] Juckins, Christopher. "Begin your Search of Past Amtrak Performance." *Amtrak Status Maps Archive Database*, https://juckins.net/amtrak_status/archive/html/home.php.

[17] Juckins, Christopher. "Browse Old Amtrak Timetables and Schedules." *Amtrak Status Maps Archive Database*, https://juckins.net/amtrak_timetables/archive/home.php.

[18] Juckins, Christopher. "Frequently Asked Questions." *Amtrak Status Maps Archive Database*, https://juckins.net/amtrak_status/archive/html/faq.php.

[19] Juckins, Christopher. "History for Amtrak 3 originating 01/23/2021 (Sa)." *Amtrak Status Maps Archive Database*, https://juckins.net/amtrak_status/archive/html/history_file.php?train_num=3&date_start=01/23/2021&date_end=01/23/2021&dfon=1&df1=1&df2=1&df3=1&df4=1&df5=1&df6=1&df7=1&sd=-1&c=-1.

[20] Juckins, Christopher. "History for Amtrak 5 originating 06/26/2015 (Fr)." *Amtrak Status Maps Archive Database*, https://juckins.net/amtrak_status/archive/html/history_file.php?train_num=5&date_start=06/26/2015&date_end=06/26/2015&dfon=1&df1=1&df2=1&df3=1&df4=1&df5=1&df6=1&df7=1&sd=-1&c=-1.

[21] Juckins, Christopher. "History for Amtrak 52 originating 05/31/2007 (Th)." *Amtrak Status Maps Archive Database*, https://juckins.net/amtrak_status/archive/html/history_file.php?train_num=52&date_start=05/31/2007&date_end=05/31/2007&dfon=1&df1=1&df2=1&df3=1&df4=1&df5=1&df6=1&df7=1&sd=-1&c=-1.

[22] Juckins, Christopher. "History for Amtrak 53 Originating 01/11/2023 (We)." *Amtrak Status Maps Archive Database*, https://juckins.net/amtrak_status/archive/html/history_file.php?train_num=53&date_start=01%2F11%2F2023&date_end=01%2F11%2F2023&df1=1&df2=1&df3=1&df4=1&df5=1&df6=1&df7=1&sort=origin_date&sort_dir=DESC&sd=-1&oper=AND&c=-1&dfon=1.

[23] Juckins, Christopher. "History for Amtrak 8 originating 03/17/2017 (Fr)." *Amtrak Status Maps Archive Database*, https://juckins.net/amtrak_status/archive/html/history_file.php?train_num=8&date_start=03/17/2017&date_end=03/17/2017&dfon=1&df1=1&df2=1&df3=1&df4=1&df5=1&df6=1&df7=1&sd=-1&c=-1.

[24] Juckins, Christopher. "Past Amtrak Delays - Average or Median Values by Train Number for Each Station." *Amtrak Status Maps Archive Database*,

https://juckins.net/amtrak_status/archive/html/average_delays.php.

[25] Kelly, Bruce E. "Idaho Weather Disrupts BNSF, up, MRL and Amtrak Services." *Railway Age*, Simmons-Boardman Publishing Inc., 20 Mar. 2017, https://www.railwayage.com/regulatory/idaho-spring-brings-woes-for-bnsf-up-and-mrl/.

[26] McDaniel, Michael. "None Injured after Train Derails near Holbrook." *Azcentral*, Www.azcentral.com, 24 Jan. 2021, https://www.azcentral.com/story/news/local/arizona-breaking/2021/01/24/train-derails-near-joseph-city-holbrook-no-injuries-danger/6696994002/.

[27] "Metrics and Minimum Standards for Intercity Passenger Rail Service." *Federal Railroad Administration*, U.S. Department Of Transportation, 16 Nov. 2020, https://railroads.dot.gov/legislation-regulations/regulations-rulemaking/metrics-and-minimum-standards-intercity-passenger.

[28] Newsdesk, KTVO. "Des Moines River expected to rise above flood stage." *KTVO*, 24 June 2015, https://ktvo.com/news/local/des-moines-river-expected-to-rise-above-flood-stage?id=1222117.

[29] "The Northeast Corridor." *Northeast Corridor Commission*, https://nec-commission.com/corridor/#:~:text=The%20Northeast%20Corridor&text=The%20Northeast%20Corridor%20main%20line,rail%20as%20well%20as%20freight.

[30] ParanoidAndroid. "Some Derailments, Landslides & 40-Hour Late #8 (17)." *Amtrak*

Unlimited, XenForo, 20 Mar. 2017, https://www.amtraktrains.com/threads/some-derailments-landslides-40-hour-late-8-17.70192/.

[31] Pascale, Jordan. "Virginia Finalizes $3.7 Billion Deal To Acquire Tracks From CSX." *NPR*, NPR, 31 Mar. 2021, https://www.npr.org/local/305/2021/03/31/983044429/virginia-finalizes-3-7-billion-deal-to-acquire-tracks-from-c-s-x.

[32] "Quarterly Report on the Performance and Service Quality of Intercity Passenger Train Operations." *Federal Railroad Administration*, U.S. Department Of Transportation, 27 Oct. 2022, https://railroads.dot.gov/sites/fra.dot.gov/files/2023-02/DOT_FRA_MetricsStandards_FY22Q4_022323_PDFa.pdf.

[33] "Rising from the Ashes: Dry Canyon Bridge Rebuilt in One Month." *InsideTrack*, Union Pacific, 5 Aug. 2021, https://www.up.com/aboutup/community/inside_track/dry-canyon-bridge-repens-early.htm.

[34] Siess, Joe. "Amtrak Closes Travel from Eugene to Sacramento Because of California Wildfire Damage." *The Register-Guard*, Www.registerguard.com, 12 July 2021, https://www.registerguard.com/story/news/2021/07/12/amtrak-closes-travel-eugene-sacramento-because-wildfire-damage/7940675002/.

[35] Slijm, JW. "Weather, Derailment Freighttrain and Amount of Freighttraffic as a

Result of That. You Can Follow It with This Link below. Press F5 for New Refresh Information. The Train Is Going to a Delay of 39 Hours. > Https://Asm.transitdocs.com/Train/2021/1/23/3." *Twitter*, 26 Jan. 2021, https://twitter.com/eurovermeer/status/1354162010452336641.

[36] "STATE OF VIRGINIA." *Amtrak*, https://www.amtrak.com/content/dam/projects/dotcom/english/public/documents/corporate/statefactsheets/VIRGINIA17.pdf.

[37] Sweeney, Sam, and Meredith Deliso. "Delayed Amtrak Auto Train Arrives 37 Hours after Departure." *ABC News,* ABC News Network, 11 Jan. 2023, https://abcnews.go.com/US/passengers-stuck-amtrak-auto-train-after-freight-derailment/story?id=96348439.

[38] Willamette Valley Railfanning. "Today's Amtrak '1014', Klamath Falls to Seattle Temporary Train Leaving Salem, OR over an Hour Late. Fires Have Destroyed the UP Line in Norther." *Facebook*, 21 July 2021, https://www.facebook.com/Oregonrailfan/videos/todays-amtrak-1014-klamath-falls-to-seattle-temporary-train-leaving-salem-or-ove/5736305846440777/.